NASA Conference Publication 3063

Joint University Program for Air Transportation Research 1988–1989

1988–1989 year-end summary

NASA

NASA Conference Publication 3063

Joint University Program for Air Transportation Research 1988–1989

Compiled by
Frederick R. Morrell
Langley Research Center
Hampton, Virginia

1988–1989 year-end summary of a program sponsored by the National Aeronautics and Space Administration and the Federal Aviation Administration

NASA
National Aeronautics and Space Administration
Office of Management
Scientific and Technical Information Division
1990

PREFACE

The Joint University Program (JUP) for Air Transportation Research is a coordinated set of three grants sponsored by NASA Langley Research Center and the Federal Aviation Administration (FAA), one each with the Massachusetts Institute of Technology (NGL-22-009-640), Ohio University (NGR-36-009-017), and Princeton University (NGL-31-001-252). These research grants, which were instituted in 1971, build on the strengths of each institution. The goals of this program are consistent with the aeronautical interests of both NASA and the FAA in furthering the safety and efficiency of the National Airspace System. The continued development of the National Airspace System, however, requires advanced technology from a variety of disciplines, especially in the areas of computer science, guidance and control theory and practice, aircraft performance, flight dynamics, and applied experimental psychology. The Joint University Program was created to provide new methods for interdisciplinary education to develop research workers to solve these large-scale problems. Each university, which submits a separate proposal yearly, is dealt with individually by NASA and the FAA. At the completion of each research task, a comprehensive and detailed report is issued for distribution to the program participants. Typically, this is a thesis that fulfills the requirements for an advanced degree or a report describing an undergraduate research project. Also, papers are submitted to technical conferences and archival journals. These papers serve the JUP as visibility to national and international audiences.

The periodic reviews held at the schools and at a NASA or FAA facility are an important feature of the program. At these reviews, the program participants, both graduate and undergraduate, have an opportunity to present their research activities to their peers, to professors, and to invited guests from government and industry.

This publication represents the ninth in the series of yearly summaries of the program. (The 1987 summary appears in NASA CP-3028.) Most of the material is the effort of students supported by the research grants.

Four types of contributions are included in this publication: a summary of ongoing research relevant to the Joint University Program is presented by each principal investigator, completed works are represented by full technical papers, research previously in the open literature (e.g., theses or journal articles) is presented in an annotated bibliography, and status reports of ongoing research are represented by copies of presentations with accompanying text.

The use of trade names or manufacturers in this report does not constitute an official endorsement of such products or manufacturers, either expressed or implied, by the National Aeronautics and Space Administration or the Federal Aviation Administration.

Frederick R. Morrell
NASA Langley Research Center

CONTENTS

PRINCETON UNIVERSITY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY 1988-1989

Robert W. Simpson
Flight Transportation Laboratory (FTL)
Massachusetts Institute of Technology
Cambridge, MA

1. INTRODUCTION

There are four areas of research being pursued in 1988 under sponsorship of the FAA/NASA Joint University Research Program, and one area which has been completed. The four active areas were:

- Automatic Speech Recognition in Air Traffic Control
- A Rule-Based Planning and Scheduling System
- Modeling of Ice Accretion on Aircraft in Glaze Icing Conditions
- Cockpit Display of Hazardous Weather Information

The completed project was:

- Terminal Area Flight Generation Using Parallel Constraint Propagation (Sadoune, FTL Report R89-1)

2. AUTOMATIC SPEECH RECOGNITION (ASR) IN AIR TRAFFIC CONTROL

Today, the Air Traffic Control system relies on verbal communication between the air traffic controllers and the pilots of the aircraft in the controlled airspace. Although a computer system exists that processes radar and other information regarding the aircraft, the information contained within the verbal communications is not retained. The introduction of ASR technology would allow this information to be captured. The purpose of this research effort is to demonstrate the feasibility of using ASR technology within the ATC environment and to address the problems involved, especially the relevant human factors issues. Off-the-shelf ASR technology will be used in conjunction with FTL's real-time ATC simulator running on the laboratory's TI-Explorer Lisp machines.

At present, new hardware is being acquired. A paper presented by Joakim Karlsson on this topic at the Ohio University Research Progress Review Meeting in June 1989 is included in this report.

3. A RULE-BASED PLANNING AND SCHEDULING SYSTEM

Planning denotes the formulation of a detailed scheme, program, or method worked out beforehand for the accomplishment of a goal. It involves the analysis of the desired goal and its division into sub-goals which are subsequently treated in the same way until a set of primitive objectives is obtained. A rational plan is prepared by a reasoner for execution by one or more actors or agents who perform actions to achieve the objectives. A reasoner is a cognitive system (human or machine) capable of some level of logical deliberation.

Computers have been employed to assist in the creation of plans almost from their inception. One of the first uses of computers, shortly after World War II, was to solve large linear optimization problems for military planners. However, the actual creation of a plan by a computer did not occur for another thirty years. The reason for this delay is that planning is a cognitive process requiring symbol manipulation, rather than strictly computation. The creation of plans by a computer had to await the development of a mature symbol manipulation language, such as *Lisp*.

A paper presented by Lyman Hazelton Jr. discussing the logical components of a rule-based planning and scheduling system presented at the MIT Research Progress Review Meeting in October 1989 is included in this report.

4. MODELING OF ICE ACCRETION ON AIRCRAFT IN GLAZE ICING CONDITIONS

The work in aircraft icing over the past year has focused on the fundamental aspects of glaze ice accretion, with the goal of improving analytical ice accretion models. Current ice accretion models do not perform well within the glaze ice regime. Glaze icing is characterized by rough and irregular ice accretions and occurs at relatively warm temperatures, i.e., a few degrees below freezing. In this regime the ice accretion is controlled by the removal of latent heat by convective heat transfer. This heat transfer is strongly dependent on the roughness of

the ice surface. Current ice accretion models assume a uniform ice surface roughness that is an input parameter in the models. Over the past year, studies have been conducted on the generation of surface roughness on accreting ice surfaces with the goal of providing a deterministic surface roughness in the ice accretion models.

A series of icing wind tunnel tests were conducted at the NASA Lewis Icing Research Tunnel, the B.F. Goodrich Icing Test Facility and the Data Products of New England Icing Test Facility. The ice accretion and surface roughness were observed on simple circular cylinders by visible and infrared photography. Based on these tests, a relatively simple modification to the existing ice accretion models has been proposed. Instead of a single uniform surface roughness, several zones of variable surface roughness are included to correspond with experimental observations. The Multi-Zone model has been implemented in the LEWICE ice accretion code in a preliminary manner and has shown a significant improvement in ice shape prediction in the glaze ice regime. In addition, the experimental and analytical efforts have focused on the need for correlations between the physical surface roughness and the equivalent sand grain roughness which is commonly used in the analytical models.

5. COCKPIT DISPLAY OF HAZARDOUS WEATHER INFORMATION

Information transfer and display issues associated with the dissemination of hazardous weather warnings are studied in the context of windshear alerts. The July 11, 1988 microburst events in which several air carrier aircraft penetrated microbursts during the Denver Terminal Doppler Weather Radar (TDWR) operational evaluation were analyzed in terms of information transfer and the effectiveness of the microburst alerts. Information transfer, message content and display issues associated with microburst alerts generated from ground-based sources were evaluated by means of air carrier pilot opinion surveys and part task simulator studies. A paper presented by Craig Wanke and John Hansman at the Ohio University Research Progress Review Meeting in June 1989 is included in this report.

ANNOTATED REFERENCE OF 1988 PUBLICATION

1. Sadoune, M.M., Terminal Area Flight Path Generation Using Parallel Constraint Propagation, FTL Report R89-1, May 1989, Flight Transportation Laboratory, MIT, Cambridge MA 02139.

A Flight Path Generator is defined as the module of an automated Air Traffic Control system which plans aircraft trajectories in the terminal area with respect to operational constraints. The flight path plans have to be feasible and must not violate separation criteria.

The problem of terminal area trajectory planning is structured by putting the emphasis on knowledge representation and air-space organization. A well-defined and expressive semantics relying on the use of flexible patterns is designed to represent aircraft motion and flight paths. These patterns are defined so as to minimize the need for replanning and to smoothly accommodate operational deviations.

Flight paths are specified by an accumulation of constraints. A parallel, asynchronous implementation of a computational model based on the propagation of constraints provides mechanisms to efficiently build feasible flight path plans.

A methodology for a fast and robust conflict detection between flight path plans is introduced. It is based on a cascaded filtering of the stream of feasible flight paths and combines the benefits of a symbolic representation and of numerical computation with a high degree of parallelism.

The Flight Path Generator is designed with the goal of implementing a portable and evolving tool which could be inserted in controllers' routine with a minimum disruption of present procedures. Eventually, a model of aircraft interaction provides a framework to rethink the notion of Separation Standards.

ANNOTATED BIBLIOGRAPHY OF 1988 PUBLICATIONS

A. Hansman, R.J., Yamaguchi, K., Berkowitz, B., and Potapczuk, M., Modeling of Surface Roughness Effects on Glaze Ice Accretion, AIAA-89-0734, AIAA Aerospace Sciences, January 1989.

Experimental observations from several icing test facilities of ice surface roughness are described. A Multi-Zone surface roughness model is proposed and implemented. Significant improvement in glaze ice accretion modeling for several cases is demonstrated.

B. Hansman, R.J., Kirby, M.S., and Lichtenfels, F., Ultrasonic Techniques for Aircraft Ice Accretion Measurement, AIAA-88-4656CP, AIAA/NASA/AFWAL Conferences on Sensors and Measurement Techniques for Aeronautical Applications, September 1988.

Ultrasonic techniques for aircraft ice accretion measurement are described. Wind tunnel evaluations in the NASA Icing Research Tunnel and flight test evaluations on the NASA Icing Research Aircraft and the Boeing 757 test aircraft are discussed.

C. Hansman, R.J., and Turnock, S.R., Investigation of Surface Water Behavior During Glaze Ice Accretion, Journal of Aircraft, Vol. 26, No. 2, February 1989.

Investigations of the behavior of surface water on accreting ice surfaces are presented. The surface water behavior controls the surface roughness which subsequently influences the convective heat transfer and the glaze ice accretion where heat transfer is the limiting mechanism for ice accretion.

D. Hansman, R.J. and Wanke C., Cockpit Display of Hazardous Weather Information, AIAA-89-0808, AIAA Aerospace Sciences Meeting, January 1989.

Information transfer issues associated with hazardous weather alerts are studied in the context of microburst warnings. Case studies, pilot opinion surveys and part task simulator studies are presented.

E. Hansman, R.J., The Influence of Ice Accretion Physics on the Forecasting of Aircraft Icing Conditions, Paper 5-1, Proceedings of the Third International Conference on the Aviation Weather System, January 1989.

The detailed physics of aircraft ice accretion processes are presented in terms of meteorological variables commonly used to forecast icing conditions. The nonlinear dependence of icing severity on these variables is discussed as well as the consequences for icing severity forecasting.

Automatic Speech Recognition in Air Traffic Control

Joakim Karlsson
Flight Transportation Laboratory
Massachusetts Institute of Technology
Cambridge, MA

Automatic Speech Recognition (ASR) technology and its application to the Air Traffic Control system are described. The advantages of applying ASR to Air Traffic Control, as well as criteria for choosing a suitable ASR system, are presented. Results from previous research and directions for future work at the Flight Transportation Laboratory are outlined.

Introduction

M.I.T.'s Flight Transportation Laboratory (FTL) is renewing its research on the application of Automatic Speech Recognition (ASR) technology to Air Traffic Control (ATC). This report presents an overview of the available technology and its potential use within the ATC system. ATC is a suitable candidate for the application of speech input/output technology due to the well-defined syntax and existing reliance on voice communication. Other motivations for introducing ASR into the Air Traffic Control environment are listed within the body of this report. Furthermore, past research efforts are described, with emphasis on work already completed by the Flight Transportation Laboratory. Finally, directions for future research are outlined.

- **Just what is Automatic Speech Recognition (ASR) anyway?**
- **ASR in Air Traffic Control.**
- **Some motivations for using ASR in Air Trafic Control.**
- **Previous work.**
- **Conclusions from Trikas' work.**
- **Work to be done at the Flight Transportation Laboratory.**

Automatic Speech Recognition

ASR systems consist of hardware and software that convert verbal input into machine-useable form (i.e., "text"). These systems can be categorized by three basic parameters: Speaker dependence/independence describes whether the system has to be trained by the user before operational use (speaker dependent), or whether it can be used by any user without specific training (speaker independent). Discrete/connected/continuous speech recognition describes the extent to which naturally spoken speech can be recognized. Single-utterance (isolated-speech) recognizers impose severe constraints on the user's manner of speech, but are relatively easy to implement. Connected speech recognizers allow the user to speak at a normal rate, but finite pauses must be inserted between each word. A continuous speech system recognizes input spoken at a natural rate, with no artificial pauses. Finally, the number of words that the system can recognize at any time (active vocabulary size) is a critical application and performance parameter.

An Automatic Speech Recognition (ASR) system is a system that recognizes verbal input and translates it into text. There are three basic factors that categorize an ASR system:

- **Speaker dependence/independence.**
- **Discrete, connected, or continuous speech recognition.**
- **Vocabulary size.**

ASR in Air Traffic Control

Today, the Air Traffic Control system relies on verbal communication between the air traffic controllers and the pilots of the aircraft in the controlled airspace. Although a computer system exists that processes radar and other information regarding the aircraft, the information contained within the verbal communications is not retained. The introduction of ASR technology would allow this information to be captured. The purpose of this research effort is to demonstrate the feasibility of using ASR technology within the ATC environment and to address the problems involved, especially the relevant human factors issues. Off-the-shelf ASR technology will be used in conjunction with FTL's real-time ATC simulator running on the laboratory's TI-Explorer Lisp machines.

We want the "computer" to capture the information given by the controller to aircraft, so that it can be processed. In this particular project, we want to start by using ASR to drive the Flight Transportation Laboratory's real-time ATC simulator.

Why Use ASR in ATC

There are several strong motivations for introducing speech input/output technology into the Air Traffic Control system. Communications are already in the verbal form, and the syntax used is clearly defined by the FAA and has to some degree been designed to reduce the possibility of communication errors. The use of voice as an input modality allows for a high information throughput capacity and allows the controllers to keep their eyes and hands busy controlling traffic. Once the verbal information has been captured, it can be transferred to the aircraft via Mode S, conformance monitoring can be improved, and routine clearances can be pre-stored during less busy periods.

- **ATC communication is verbal.**
- **ATC syntax is clearly defined.**
- **ATC training can be automated.**
- **High information throughput.**
- **ASR allows controller to use hands and eyes where they belong.**
- **Captured information can be transmitted to aircraft via Mode S.**
- **Conflict alert can be improved.**
- **Clearances can be pre-stored.**

Previous Research

ASR technology can be used in many aviation and non-aviation applications, and as a result, much research has been conducted on the use of speech input/output in general. However, relatively little research has been dedicated toward the application of ASR to Air Traffic Control. The research to be undertaken within the framework of this project will be a continuation of the initial work presented in Thanassis Trikas' S.M. thesis, *Automated Speech Recognition in Air Traffic Control* (FTL report R87-2).

A lot of research has been done on ASR, but not much in conjunction with ATC:

- **FTL: Thanassis Trikas S.M. work.**
- **Arthur Gerstenfeld (Worcester Polytechnic Institute/UFA, Inc.): Emphasis on ATC training.**
- **ITT Defense Communications Division VRS 1280 demonstration.**

Trikas' Conclusions

Trikas' thesis demonstrated the feasibility of using ASR technology in conjunction with an ATC simulator, utilizing a relatively small vocabulary. An initial error correction strategy based on verbal correction commands alone proved to be unacceptable. Also, problems related to speech articulation variations were encountered. In the process of evaluating his experiment, Trikas implicitly set forth a set of criteria for selecting a suitable ASR system.

Trikas' S.M. thesis was essentially a proof of concept of using ASR in ATC:

- **ASR <u>can</u> be used with the ATC simulator (with an active vocabulary of only 64 words).**
- **Correction of recognition errors using voice alone is not feasible.**
- **Problems with sensitivity to variations in articulation.**
- **Developed criteria for choosing a suitable ASR system.**

Selecting the Right ASR System

The first step in renewing FTL's ASR research effort will be to select a suitable hardware system. For this purpose, performance criteria specific to ATC applications of speech input/output technology have been defined.

Our particular application calls for the following ASR requirements:

- **Speaker independence <u>not</u> required.**
- **Continuous speech recognition.**
- **Vocabulary size 200-300 words.**
- **95% baseline recognition accuracy.**
- **Well-designed training procedure.**
- **Open architecture.**
- **Reduced sensitivity to variations.**
- **Short recognition delays (1-4 s).**

Future Work

The future research to be conducted at FTL will be based on previous work completed by Trikas. Hence, his system setup must be reactivated. In order to improve the simulation and the overall performance of the system, new hardware will be acquired. The actual research will concentrate on the introduction of multi-modal input, improved error correction and recognition accuracy, the evaluation of Mode S usage, and the application of ASR to secondary functions.

- **Reassemble Trikas' system.**
- **Evaluate current ASR technology.**
- **Acquire a new ASR system.**
- **Introduce multi-modal input.**
- **Increase number of commands and responses to improve simulation.**
- **Improve error checking/correction, as well as recognition accuracy.**
- **Evaluate Mode S usage.**
- **Use ASR for functions other than ATC commands.**

COCKPIT DISPLAY OF HAZARDOUS WIND SHEAR INFORMATION

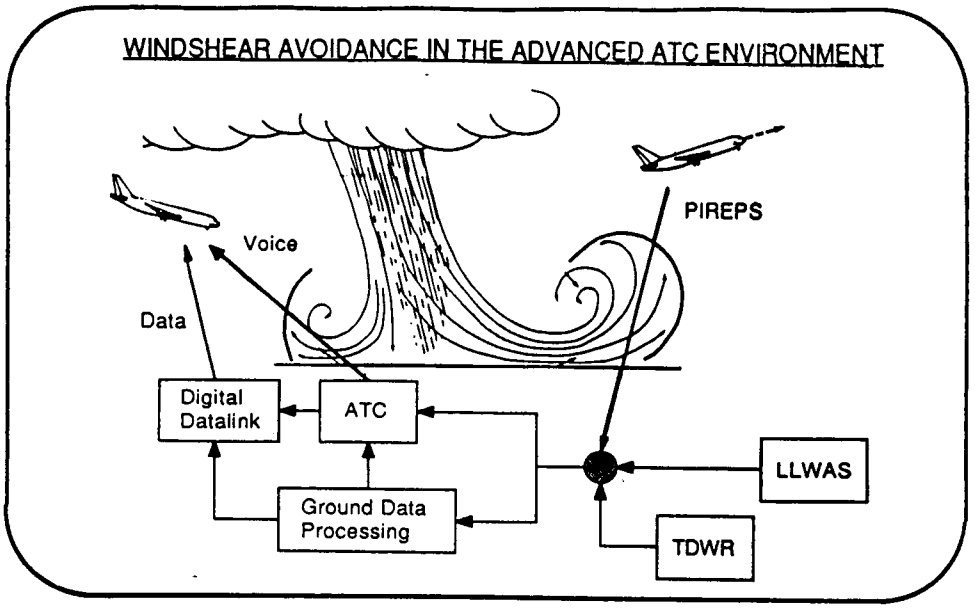
Craig Wanke
R. John Hansman
Massachusetts Institute of Technology

Based on the current status of windshear sensors and candidate data dissemination systems, the near-term capabilities for windshear avoidance will most likely include:

- Ground-based detection:
  - TDWR (Terminal Doppler Weather Radar)
  - LLWAS (Low-Level Windshear Alert System)
  - Automated PIREPS
- Ground-Air datalinks:
  - ATC voice channels
  - Mode-S digital datalink
  - ACARS alphanumeric datalink

The possible datapaths for integration of these systems are illustrated in the diagram.

In the future, airborne windshear detection systems such as lidars, passive IR detectors, or airborne Doppler radars may also become available. Possible future datalinks include satellite downlink and specialized en route weather channels.

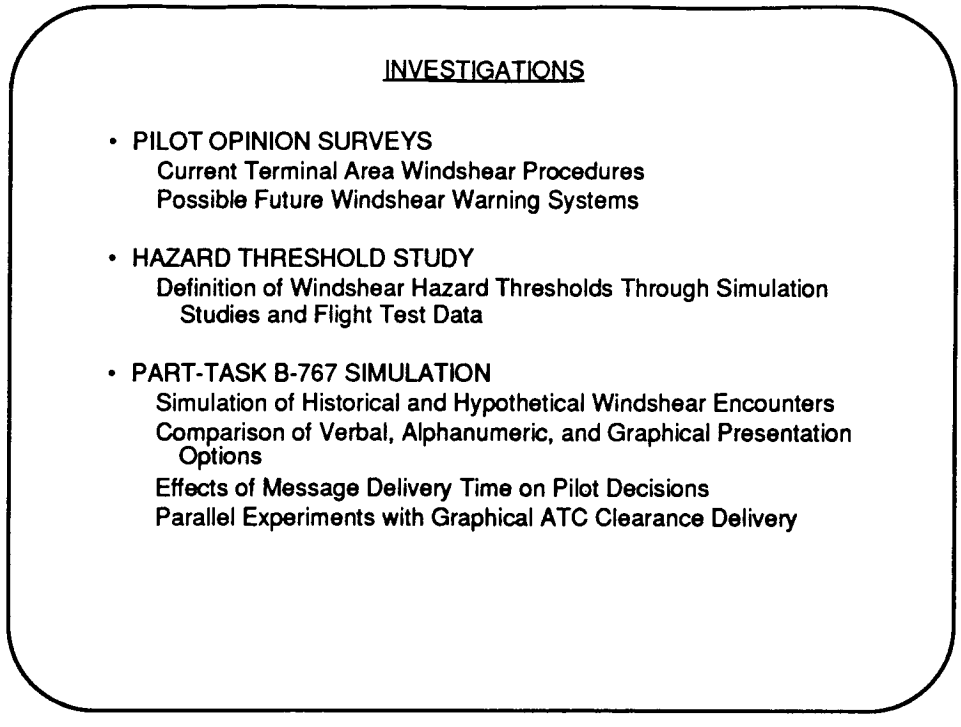


Uplink of ground-measured windshear information to the flight crew presents a number of problems. Among these are:

- Hazard Assessment:
  Based on ground-measured information and the current tactical situation, what constitutes a hazardous windshear?

- Information Issues:
  What to send: what is 'critical'
  When to send it
  How to send it
  How to present critical info to the flight crew

The MIT effort is using the three methods outlined to address these issues.



The survey has two sections: Current Procedures and Future Windshear Warning Systems. The first part included questions about:

- Pilot impressions of the hazard posed by microbursts
- Pilot confidence in windshear information obtained from ATIS, LLWAS, and PIREPS
- Pilot evaluation of the adequacy of currently available windshear alert data and the need for better and more timely alerts

The section on future procedures included questions about

- Modes of information relay and presentation
- When the crew should be alerted
- How much head-to-tail windshear is a threat
- What items of windshear information are important
- Who makes the avoidance decision - pilot or controller

PILOT OPINION SURVEYS

- Obtain flight crew evaluations of current windshear warning and avoidance systems and procedures
- Obtain flight crew feedback on future windshear warning systems and possible display formats
- Distribution: 250 United A/L flight crews
- Current Data Set: 47
  51% of respondents have had a hazardous windshear encounter
- Data are being used to design part-task simulator experiments with advanced graphic and alphanumeric display formats

General results indicate that microbursts are perceived as a major safety threat and that currently available windshear alert data are not sufficient for safe terminal area operations. Also, there is strong pilot support for an improved windshear alerting system. The fairly low windshear threshold specified by the respondents is another indication of the belief that windshear is a major threat.

Sample comments from the last question listed:

"Only the pilot can judge the capabilities of his a/c type and his ability to cope with reported data."

"(Controller) Must make judgment calls to report only those shears meeting some minimum standards to avoid reporting 'everything'."

51% of the respondents have had (by their own definition) a hazardous windshear encounter.

PILOT SURVEYS: GENERAL RESULTS

	Agree	Disagree
Microbursts pose a major safety hazard to transport category aircraft.	90.2%	4.9%
Currently available windshear alert data are sufficient for safe operation.	14.6%	43.9%
A system to provide crews better and more timely windshear information is necessary.	97.6%	0.0%

• Perceived windshear warning threshold:	Advisory: 10.6 kts Warning: 15.1 kts
• Who should have the responsibility for judging the threat due to a particular windshear event from the (assumed reliable) available data?	PILOT: 83.0% CONTROLLER: 9.5%

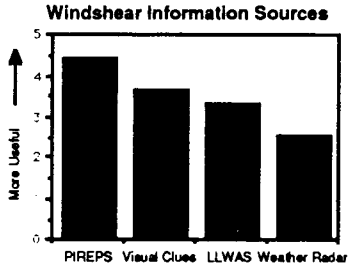
The data are from the question:

Listed below are four currently available sources of information about windshear in the terminal area. Please rank them in order of usefulness, from 1 (most useful) to 4 (least useful).

_____ Low Level Windshear Alert System (LLWAS)
_____ Pilot Reports (PIREPS)
_____ Airborne Weather Radar
_____ Visual Clues (Thunderstorms, Virga etc.)

Pilot Rankings: Usefulness for Windshear Avoidance of

Pilot Reports
Visual Clues
LLWAS (Low Level Windshear Alert System)
Airborne Weather Radar



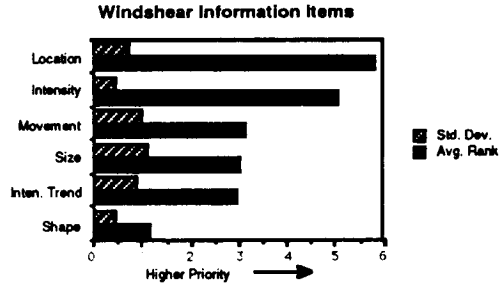
The data are from the question:

A windshear alert could contain the following items of information. Please rank them in order of importance. (1 = most important, 6 = least important)

_____ Location
_____ Intensity
_____ Size
_____ Shape
_____ Intensity Trend
_____ Movement

Pilot Rankings: Usefulness of Available Windshear Data

Location	Size
Intensity	Intensity Trend
Movement	Shape



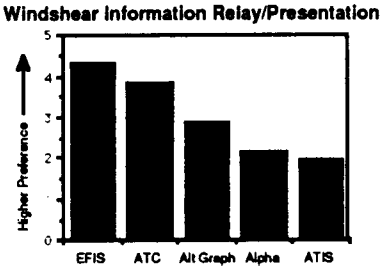
The data are from the question:

Assuming windshear is detected by reliable ground-based sensors, how should this information be relayed to the flight deck? Please rank in order of preference. (1 = most preferable, 5 = least preferable)

_____ Voice (ATIS)
_____ Voice (ATC)
_____ Alphanumeric/Text uplink (similar to ACARS)
_____ Graphical display of windshear location on EFIS display
_____ Graphical display of windshear location on separate graphic device

Pilot Rankings: Mode of data relay/presentation:

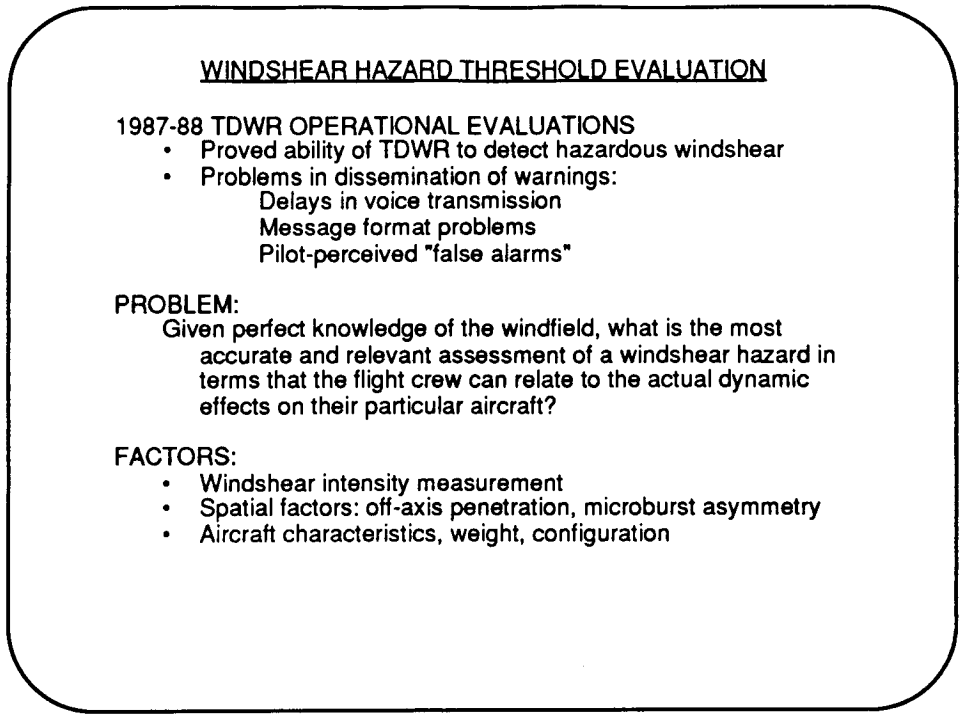
Verbal (ATC)
EFIS EHSI (Moving Map) Display
Alternate Graphical Display
Alphanumeric Display
On ATIS



The hazard assessment part of this project is motivated by the TDWR (Terminal Doppler Weather Radar) operational evaluations performed by MIT Lincoln Laboratory and NCAR over the past three years. In these experiments, TDWR proved able to detect hazardous microbursts nearly 100% of the time with a false alarm rate of 5%.

Despite the accuracy of the radar, some dangerous incidents still occurred. One potential problem is assessment of the hazard due to a particular microburst in a particular situation. The hazard reported by the detection system must relate directly to what the threatened aircraft would experience if it does penetrate the microburst. If a large threat is predicted, and little dynamic effect is encountered, an accurate warning can be perceived as a 'false alarm' and damage pilot confidence in the system.

As listed, this study will compare the factors contributing to the windshear hazard and attempt to determine a useful and relevant hazard criterion.



The most critical factor in the threat evaluation is how to quantify the microburst intensity. Three possible ways are listed. The delta-V measurement is currently used by TDWR. F-factor, a measure of instantaneous threat, is more applicable to *in situ* and airborne sensor systems. An integrated energy loss calculation is another possibility, better suited to long range sensors such as TDWR.

WINDSHEAR INTENSITY MEASUREMENT

- DELTA-V

  Maximum headwind-to-tailwind component an aircraft can be expected to encounter over a specified distance

- F-FACTOR

$$= \frac{\dot{W}_x}{g} - \frac{W_h}{V}$$

  Developed for *in situ* sensors, measures loss of "potential altitude" due to immediate windfield

- INTEGRATED ENERGY LOSS

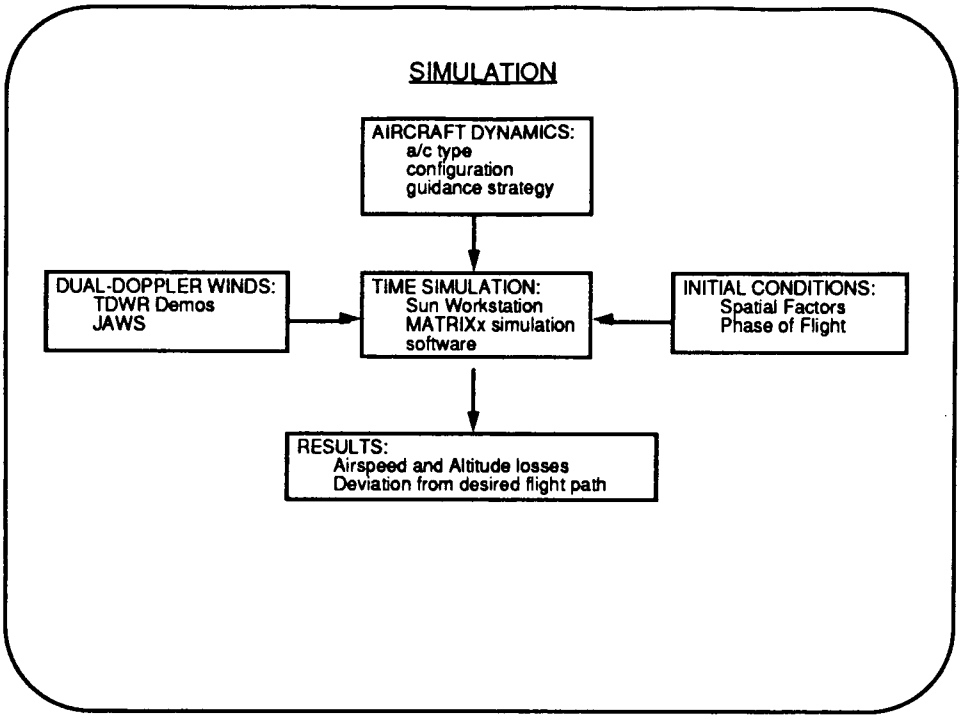
  Measures expected loss of energy over a projected flight path: including kinetic (airspeed) and potential (altitude) energy components

These issues will be evaluated through parametric flight simulations and possibly actual flight data. As listed, measured dual-Doppler windfields will be used with available aircraft models to examine the effects of various factors on the microburst hazard. Once a candidate hazard threshold has been determined, the dual-Doppler data will be degraded to simulate a single TDWR measurement and the hazard criterion will be re-evaluated. Supporting data may be gained from flight tests flown this summer by the University of North Dakota in conjunction with the TDWR evaluation.

PLANNED APPROACH

- Literature Search: Microburst-Aircraft interaction studies
- Develop Simulation
  - Aircraft Dynamics: 737-100, Cessna Citation
  - Windshear model - assume ideal sensor using
    - NCAR/MITLL dual doppler radar windfields
- Perform Parametric Studies
- Evaluate and modify hazard criteria based on studies
- Evaluate flight test data from UND Citation
- Simulate effect of sensor limitations on hazard threshold validity

The simulation is currently under development. It will be run on a Sun 3/80 workstation with MATRIXx simulation software to minimize development time and allow easy automation of parametric studies.



An ongoing focus of the MIT windshear effort is the dissemination of ground-measured and evaluated windshear information. As displayed, several information-related issues arise and are being addressed with several resources, including the previously discussed pilot opinion surveys. Current research is focused around the 767 part-task simulation.

UPLINK OF CRITICAL WINDSHEAR INFORMATION

OBJECTIVE:

Examine transfer and presentation of vital windshear information to flight crews, based on modern ground-based sensors and datalink capabilities.

ISSUES: What information to transmit
When to send information
Information mode: Verbal, Alphanumeric, Graphical

RESOURCES: Pilot surveys
Analysis of past windshear encounters (7/11/88)
GA part-task simulator study
Boeing 767 part-task simulator

On July 11, 1988 a group of very severe microbursts occurred on the final approach path to Denver/Stapleton runway 26L. Five aircraft were affected by the event. The TDWR evaluation was in progress, and successfully detected the microbursts. However, four of the five aircraft involved still attempted to land and penetrated the microbursts, in one case passing below 100 feet AGL one mile short of the runway threshold. The problem in this case was with dissemination of the warnings, not with detection of the microbursts. In this case, the effectiveness of the alert was degraded sharply by a number of factors:

- Variability of aircrew interpretation of warnings: In one case, two aircraft were given virtually identical warnings within 30 sec of each other and made entirely opposite decisions.

- Delay inherent in voice transmissions: The minimum delay between detection and message transmission was about 60 sec.

- The messages were often imbedded in a routine landing clearance message, resulting in a lack of urgency; this, combined with the high crew workload at the outer marker, may have caused the warnings to be missed or go unheeded.

- PIREPS are critical - once the lead aircraft reported windshear, later aircraft initiated earlier missed approaches.

MICROBURST EVENT: 11 JULY 1988

- Occurred during 1988 TDWR Operational Demonstration at Denver-Stapleton Airport: Doppler Radar was operational.
- 20 minute period of intense microburst activity on final approach to active runway 26L: shear intensity up to 80 kts
- 5 aircraft involved
- Microburst warnings promptly and accurately produced by TDWR system: given to final approach controller
- Despite timely warnings, 4 of 5 aircraft penetrated microbursts and encountered significant performance losses before aborting the approach
- Problems:
  - Time delays
  - Message format

This simple experiment was designed to demonstrate the gain in pilot situational awareness due to graphical displays. By linking a schematic graphical display of windshear location to a general aviation simulator, a significant improvement in pilot performance was measured. This experiment, although simple, indicates that graphical displays are desirable and that further investigation of graphical formats and measurement of achievable performance gain is required.

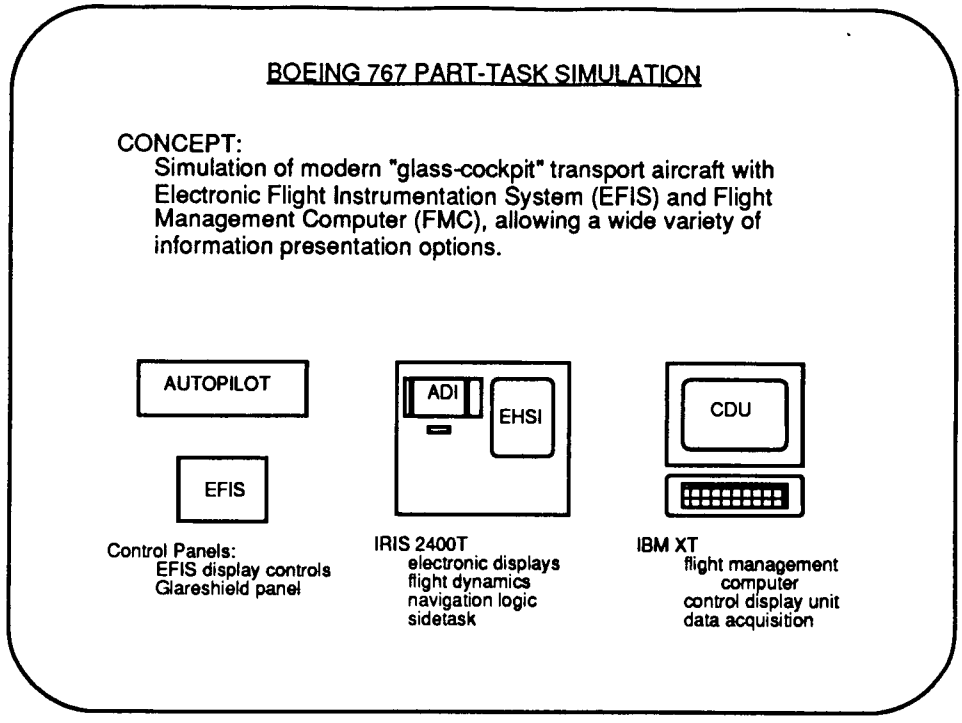
PART-TASK GA SIMULATION

- GA Simulator and IBM PC display
- Microburst occurring on final approach
- Non-standard missed approach procedure required for avoidance

Display Type Used	Avoidance Rate
Voice only	43%
Runway-fixed display without a/c position	62%
Runway-fixed display with a/c position	94%

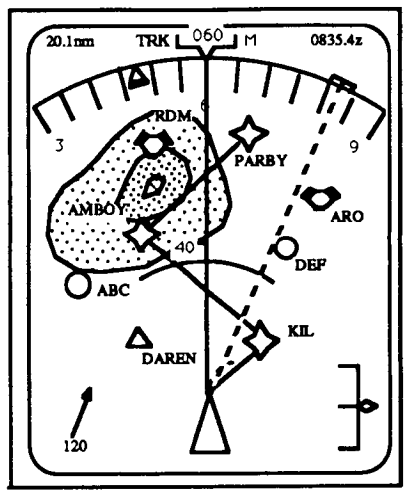
- 8 Pilots: 210 to 1700 hours total time
  16 Approaches per display

Further information presentation options are being explored through a Boeing 767 part-task simulation, which has just been completed. The 767 was selected because its advanced 'glass-cockpit' instruments allow for many possible alphanumeric or graphical presentations. The major electronic and electromechanical instruments are simulated on an IRIS graphic workstation, and an IBM is used for simulation of the major features of the Boeing FMC (flight management computer). Mockups of the autopilot control panel and electronic display control panel are linked into the system through a data acquisition unit. In addition, the crew workload is adjusted and monitored through a simple compensatory sidetask controlled by the IRIS mouse. The final component is a pair of headsets for pilot and 'controller'.



This figure illustrates a typical EHSI (Electronic Horizontal Situation Indicator) in map mode for an aircraft equipped with a flight management computer. Information displayed includes the currently programmed flight path, nearby navaids, intersections, and airports, wind information, and the aircraft's current heading and groundtrack. Weather radar returns are also overlaid on the EHSI, making it even more desirable to place windshear information on this display.

767 EHSI - MAP MODE



The initial experiments as described are designed to gather more data about modes of information presentation. The experiment is being performed in concert with a study on the benefits of delivering ATC routing amendments graphically. Each scenario has elements from both experiments, which should reduce the pilot learning curve and result in better quality data. All pilot actions, including ATC transmissions and sidetask performance, will be recorded and time-stamped by the simulation computers. Audio and video coverage will also be taken.

INITIAL SIMULATIONS

- Performed in concert with Graphical Amendment Delivery experiments
- 9 scenarios will be flown by each subject:
  - Denver-Stapleton Airport
  - 3 Initial Conditions, 3 ILS approaches
  - Modes:
    - Verbal, Alphanumeric, Graphical
  - Microburst Position:
    - Off to side of approach
    - On final approach
    - On final approach and cleared missed approach path
  - Alert times:
    - When cleared for approach
    - 2 nm outside outer marker
    - At or inside outer marker
  - Simple compensatory sidetask for workload monitoring

If 9 subjects are obtained, as described a total of 81 approaches will be flown. Allowing 3 modes of communication and 3 different microburst threat conditions, 9 points per set of parameters will be obtained. These experiments are planned for July and August of 1989.

Possible future experiments will deal with such issues as delay times and possible graphical format options.

INITIAL SIMULATIONS

SUBJECTS:

- Active Transport Category a/c crews
- Time requirement: 5 hours per subject
- Minimum 9 subjects desired for initial simulations
- Subjects are being contacted through ALPA, to be tested during BOS layovers

STATUS:

- Simulator 99% operational
- Subjects currently being sought
- Results should be available for next Tri-U Conference...

Modeling of Surface Roughness Effects on Glaze Ice Accretion * ^

R. John Hansman, Jr.* and Keiko Yamaguchi†
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, Massachusetts

Brian Berkowitz#
Sverdrup Technology, Inc.
Middleburg Heights, Ohio

and

Mark Potapczuk**
NASA Lewis Research Center
Clevland, Ohio

Abstract

A series of experimental investigations focused on studying the cause and effect of roughness on accreting glaze ice surfaces were conducted. Detailed microvideo observations were made of glaze ice accretions on 1" to 4" diameter cylinders in three icing wind tunnels (the Data Products of New England six inch test facility, the NASA Lewis Icing Research Tunnel, and the B.F. Goodrich Ice Protection Research Facility). Infrared thermal video recordings were made of accreting ice surfaces in the Goodrich facility. Distinct zones of surface water behavior were observed; a smooth wet zone in the stagnation region with a uniform water film; a rough zone where surface tension effects caused coalescence of surface water into stationary beads; a horn zone where roughness elements grow into horn shapes; a runback zone where surface water ran back as rivulets; and a dry zone where rime feathers formed. The location of the transition from the smooth to the rough zone was found to migrate with time towards the stagnation point. The behavior of the transition appeared to be controlled by boundary layer transition and bead formation mechanisms at the interface between the smooth and rough zones. Regions of wet ice growth and enhanced heat transfer were clearly visible in the infrared video recordings of glaze ice surfaces. A simple multi-zone modification to the current glaze ice accretion model was proposed to include spatial variability in surface roughness. A preliminary version of this model was implemented on the LEWICE ice accretion code and compared with experimental ice shapes. For the two cases run the multi-zone model significantly improved the prediction of the glaze ice shapes.

1.0 Introduction

Current efforts to analytically model glaze ice accretion are hampered by insufficient knowledge of the accreting ice surface roughness. In glaze icing, which normally occurs at temperatures near freezing or high liquid water contents, there is insufficient convective heat transfer to remove all of the latent heat of freezing of the impinging supercooled water droplets. The local ice accretion rate is, therefore, controlled by the local convective heat transfer. The droplet impingement behavior also plays an important role in determining where on the surface this wet growth occurs and is the controlling factor in dry or rime ice growth where there is sufficient cooling to freeze the droplets on impact.

The local convective heat transfer from a surface is known to be strongly dependent on the ice surface roughness.[1] Because of the importance of the heat transfer behavior on the ice accretion rate, the surface roughness becomes an important factor modeling glaze ice accretion. Current analytical models such as LEWICE [2] generally assume that the surface roughness is uniform and the effective sand grain roughness k is used as an input parameter in the code. In some cases ad hoc spatial heat transfer distributions have been used.[3] The magnitude of the roughness parameter k is normally determined empirically by comparison of predicted and experimental ice accretion.

The erratic performance of glaze ice accretion models and the empirical manner in which the surface roughness is treated indicates the need for a more deterministic treatment of the surface roughness. In the following, experimental investigations focused on the roughness of accreting glaze ice surfaces will be presented as well as preliminary results of a simple modification to the LEWICE code which attempts to treat the ice surface roughness in a more physical manner.

2.0 Experimental Setup For Ice Surface Observations

2.1. Experimental Facilities

Detailed observations of accreting ice surface roughness were made at several experimental icing facilities. These included the Data Products of New England small icing wind tunnel (6" X 6"), the NASA Icing Research Tunnel (9' X 7'), and the B.F. Goodrich Icing Wind Tunnel (2' X 4'). In the following these facilities will be designated as DP, IRT, and BFG respectively

2.2. Photographic Setup

Detailed photographic observations of the behavior of surface water and formation of ice roughness on cylinders during glaze ice accretion were made in the three facilities described above. Cylinder diameters from 1" to 4" were

---

* Associate Professor, Associate Fellow AIAA.
† Research Assistant, Student Member AIAA.
# Engineering Associate.
** Aerospace Engineer, Member AIAA.

*AIAA-89-0734, 27th Aerospace Sciences Meeting, 1989.
^Some original figures were not available at time of publication.

observed. Free stream velocities of 100 and 150 kts were run at liquid water contents up to 1.5 g/m^3. Tunnel total air temperatures varied from 0°C to -10°C with some colder runs made for calibration purposes. In addition, a thermal infrared (IR) video system was used during the BFG tests to observe the relationship between surface temperature and surface roughness in glaze icing conditions.

The photographic setup used in the DP and BFG tests is shown in Fig. 1. The test articles were cylinders which horizontally spanned the test section. The test section walls were either plexiglass (DP) or heated glass (BFG) to provide photographic access. Scale reference was provided by a grid mounted on a thin splitter plate at the midplane of the test section. A CCD microvideo (Camera A) with a macrolens for magnification was used to obtain a grazing angle view of the ice accretion. The camera was focused at the stagnation region near the center of the test section. An additional video camera (Camera B) was mounted slightly upstream of the cylinder to provide a view of the ice accretion looking normal to the cylinder surface. Secondary lighting was provided to obtain a clear video record of the accreting ice surface. A second grazing angle camera (Camera C) was also used in some cases to provide a less magnified view of the overall ice shape.

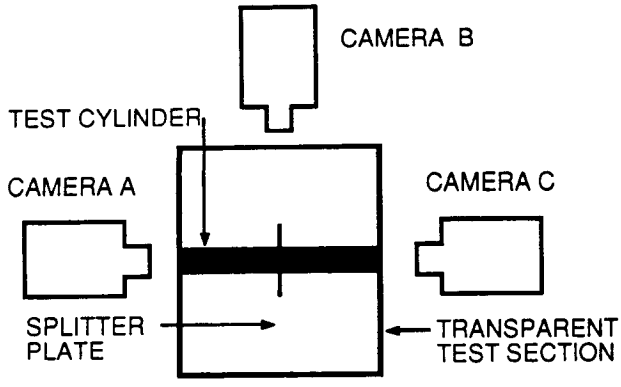


Fig. 1. Schematic view of the photographic setup for the DP and BFG icing tunnels.

The experimental setup used in the IRT tests is shown in Fig. 2. The photographic and test cylinder geometry is very similar to the DP and BFG grazing angle camera setup described above. However, in order to locate the test cylinder within the region of uniform cloud, both the test cylinder and the camera had to be moved into the tunnel. A heated waterproof fairing was provided to protect the camera. In addition, a shutter was located upstream to shield the test cylinder during initial transients in the liquid water content. Once the tunnel spray system had stabilized, the shield was removed and the cylinder began to ice. The nominal MVD used in the IRT tests was 20 microns.

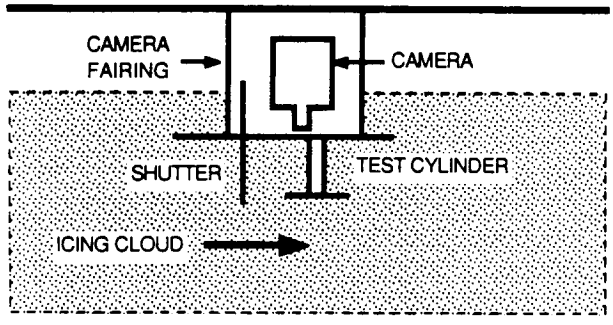


Fig. 2. Schematic view of the photographic setup in the IRT.

2.3 Infrared Video Setup

The IR thermal video system setup used in the BFG tests is shown in Fig. 3. The IR camera was mounted above the test section. Because no IR transparent windows were available, a viewing slot was cut into the top of the tunnel so that the cylinder could be viewed directly. The IR video system provided color thermographs of the accreting ice surfaces and temperature profiles along designated sections. The system was typically operated with the cooler temperature scale set to 0.5° C and a dynamic range set between the tunnel operating temperature and 0°C. Cold air blowing out of the viewing slot cooled the IR optics and caused the absolute calibration of the system to drift. The system was, therefore, calibrated "in situ" by viewing points of known temperature while the tunnel was operating.

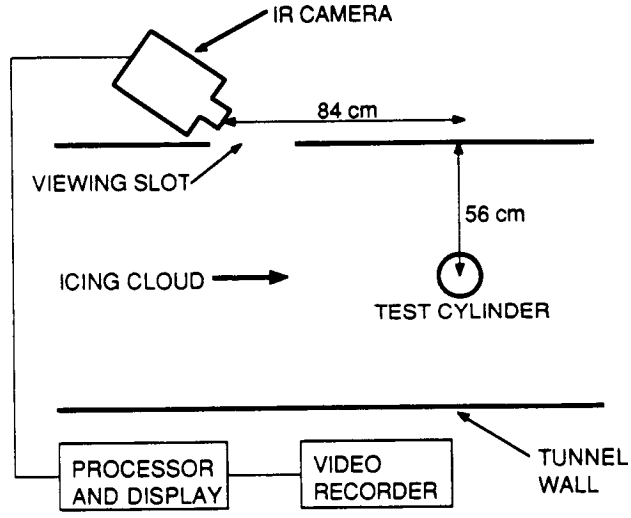


Fig. 3. Schematic view of the IR video system setup in the BFG icing tunnel.

3.0 Experimental Observations of Ice Surface Roughness

3.1 Surface Roughness Zones

During the detailed photographic analyses of accreting ice surfaces, distinct regions were observed each having a characteristic roughness and identifiable boundaries. A total of 5 roughness zones were identified although rarely were more than 3 present on any particular ice accretion. The 5 zones are described briefly below.

3.1.1 Smooth Zone

In glaze ice accretions, close to the stagnation point, the surface was observed by light reflection and IR techniques to be uniformly wet with a thin film of water at warm temperatures. The surface in this regime was smooth, with no distinctly visible roughness. The ice was translucent within the smooth zone. The extent of the smooth zone can be seen in Fig. 4 which shows a representative final glaze ice shape for a 3 min. icing exposure at 150 kts and -4.5°C at a liquid water content of 0.4 g/m^3 in the DP tunnel. Similar accretions were observed in IRT and BFG facilities.

3.1.2 Rough Zone

At some point downstream, there was a sudden transition to a significantly rougher surface. This can be seen in Fig. 4 Within this zone, there appeared to be insufficient water to maintain a uniform film. Surface tension forces dominated the water surface behavior. Runback did not occur, rather, the water tended to coalesce into the water beads first observed by Olsen, et al [4,5]. The scale length of the roughness was typically on the order of 1 mm.

Inasmuch as there was a distinct boundary between the smooth and rough zones this position could be easily identified on the grazing angle video recording. The angular position of this boundary is plotted as a function of the time in Fig. 5 for the accretion shown in Fig. 4. The transition is observed to propagate rapidly towards the stagnation region. This behavior was observed in all three test facilities (DP,IRT,BFG). The repeatable nature of the smooth-rough transition's propagation towards the stagnation region implies a clear underlying physical mechanism for the transition between the surface water behavior in the smooth and rough zones. This is discussed in more detail in Section 3.2.

The ice accretion rate was observed to be enhanced in the rough zone as compared with the smooth zone. This can be seen in the ice profile shown in Figure 4. The enhanced accretion rate is thought to be due to increased heat transfer resulting from the greater surface roughness in this zone. In all the cases observed in this study, the ice horns characteristic of glaze ice accretion were found within the rough zone.

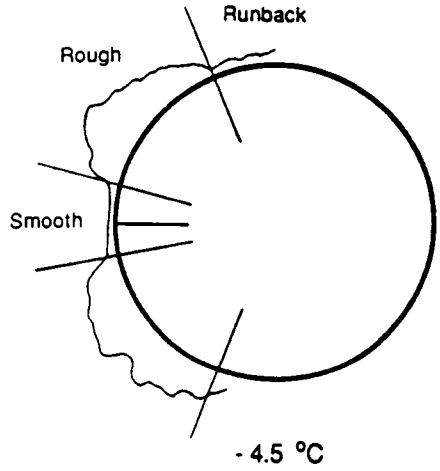


Fig. 4. Typical final glaze ice shape showing distinct roughness zones, (T=-4.5°C, W=1.0g/m^3).

3.1.3 Horn Zone

In certain glaze ice accretions particularly at high liquid water contents, some of the roughness elements within all or part of the rough zone were observed to grow into distinct protuberances. An example of such a case is shown in Fig. 6 for a very warm (-0.5°C) glaze ice accretion. Once these elements extended far enough above the ambient surface, they became essentially separate microscopic ice accretions with high collection efficiencies catching droplets which would otherwise have been swept by the accreting body. Because of the collection efficiency and the enhanced heat transfer of these protuberances they tended to grow rapidly resulting in horned ice accretions. In some warm cases such as the one shown in Fig. 6, water was observed to be torn or shed from the individual elements. This results in a water flow mechanism which is not considered in any present ice accretion models.

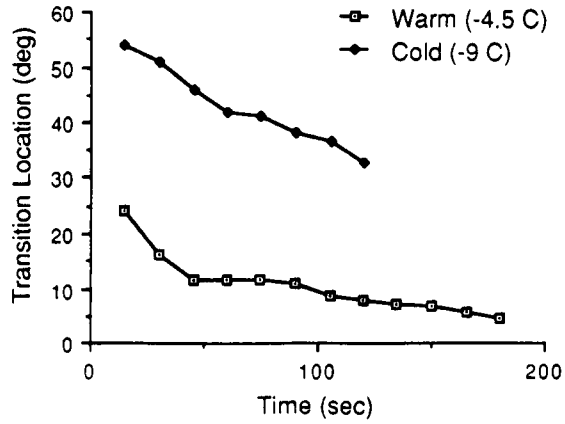


Fig. 5. Angular position versus time of smooth-rough transition location for two air temperatures.

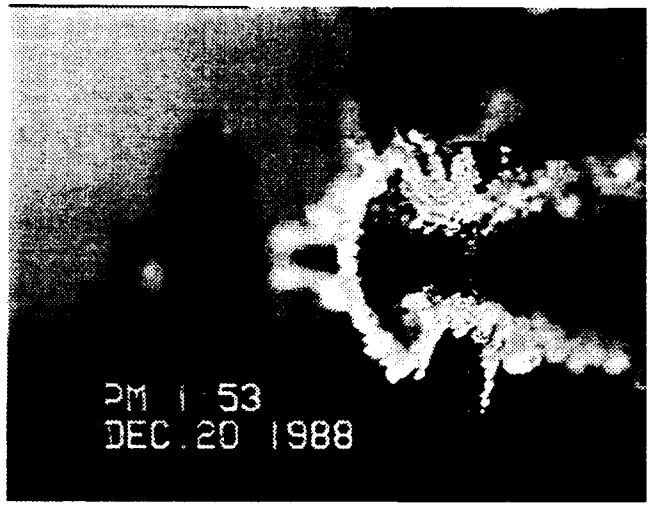


Fig. 6. Extreme example of horn elements for a warm glaze ice accretion (-0.5°C).

3.1.4 Runback Zone

At warm temperatures a runback zone was observed aft of the rough zone in the primary accretion. An example can be seen in Figure 4. This region was characterized by areas of ice interspersed with uniced surface. This ice was observed to form during an initial transition period after cloud exposure. The ice was translucent, and quite often frozen rivulets could be discerned. In warm conditions and high liquid water contents, the surface water was observed to initially runback and then stagnate at the point of flow separation. This water then slowly froze as rivulets or as large coalesced water cells. Once ice began to form in the upstream rough zone, no additional surface water was supplied to the runback zone and the ice surface remained constant.

3.1.5 Rime Feather Zone

At cold temperatures rime feathers are sometime observed to grow in the region aft of the primary accretion. The feathers were observed by the IR Video System to be dry ice accretions and propagate in the local upstream direction. It should be noted that there are intermediate conditions where neither runback or rime feathers are observed.

3.2 Effect of Icing Cloud Parameters on the Rough to Smooth Transition

The effect of icing cloud parameters on the transition between the smooth and rough surface zones was observed in an attempt to identify the underlying physical mechanisms which cause the rough surface to develop. The transition between the smooth and rough surface zones was monitored as a function of time for a variety of cloud conditions in the DP tunnel. The temperature, liquid water content and free stream velocity effects were studied individually by holding all other parameters constant at nominal values and comparing transition behavior. The nominal velocity was 150 kts, the nominal liquid water content was approximately 0.4 g/m ^{3} with a mass of 30 microns, and the nominal temperature was - 4.5°C.

The effect of cloud temperature, liquid water content and free stream velocity on the angular position of transition are shown as a function of time in Figs. 5, 7 and 8 respectively. In each case the transition point is observed to migrate towards the stagnation point as discussed above. However, a significant variation was observed in the actual transition location. It can be seen that warmer temperatures, increasing liquid water content and increasing velocity tend to move the transition point closer to the stagnation region. These observations tend to imply that increasing Reynolds number and increasing the flow of surface water cause the smooth to rough transition point to move closer to the stagnation region.

The observed Reynolds number dependance may indicate that, in certain conditions, the initial transition in surface roughness is controlled by the boundary layer. In these cases the laminar to turbulent boundary layer transition point will also be the initial smooth to rough transition point. The enhanced heat transfer in the turbulent region would cause sufficient freezing to partially dry the surface and cause bead formation. Whereas, in the laminar region, the heat transfer is low enough that the surface is coated by a uniform water film. Once the roughness zone emerges, the roughness elements themselves will cause the transition of the boundary layer if it is not fully turbulated.

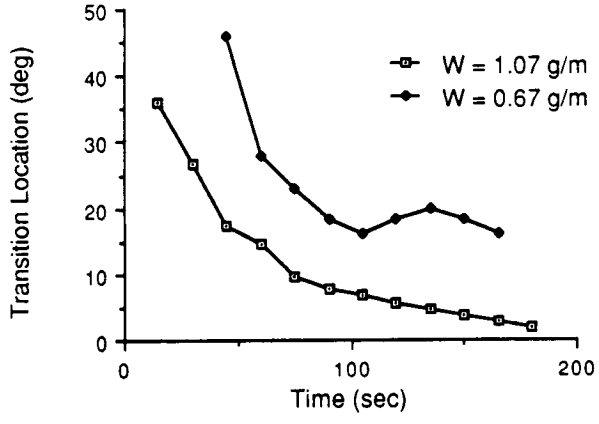


Fig. 7. Angular position versus time of smooth-rough transition location for two cloud liquid water contents.

In order to more fully evaluate this hypothesis, the initial location of the rough to smooth transition was observed as a function of Reynolds number based on cylinder diameter for a variety of icing conditions in the IRT and BFG tunnels. The results are shown in Fig. 9. While there is some scatter in the data the initial rough to smooth transition is generally observed to move towards the stagnation point with increasing Reynolds number. Also shown are the theoretical locations of the boundary layer transition based on a critical roughness element

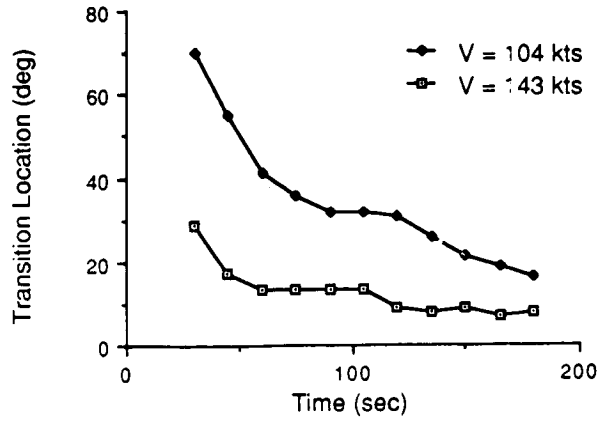


Fig. 8. Angular position of smooth-rough transition versus time for two free stream velocities.

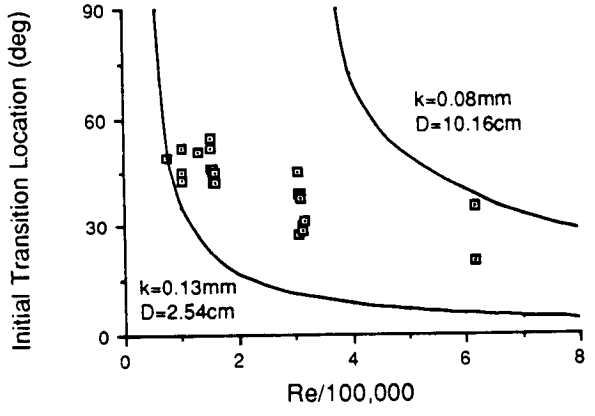


Fig. 9. Initial transition location versus Reynold's number.

Reynolds number of 600 for 0.13 mm roughness elements on a 1" cylinder and 0.08 mm roughness elements on a 4" cylinder.[6] The experimental data are bracketed by the boundary layer transition curves. Considering that the roughness in the wet stagnation region is likely to decrease with increasing Reynolds number, the experimental results do not appear to contradict the hypothesis that the initial rough to smooth transition is a result of boundary layer transition.

The influence of surface water flow on the rough to smooth transition location indicates that dynamic effects are also important. It is thought that the dynamic effects are caused by bead formation at the interface between the smooth and rough surface zones as is shown in Fig. 10. The formation of these beads at the interface causes enhanced heat transfer within the rough zone which tends to freeze out the downstream beads and dry the surface. As the surface drys, beads begin to form further upstream and the transition point will propagate towards the stagnation region as observed experimentally. By increasing the surface water flux, the rate of formation and growth of the interfacial beads is increased. This causes the observed increase in upstream propagation of the transition point with surface water flux.

Further experimental observations are under way to more fully determine the boundary layer and dynamic mechanisms which cause the transition in surface roughness.

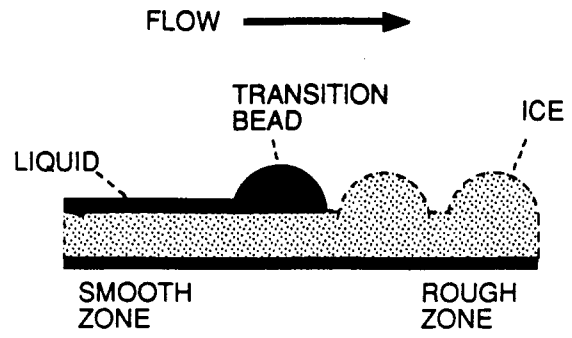


Fig. 10. Schematic representation of bead formation at the smooth to rough transition.

3.3 Infrared Observations

Typical digitized video images of a 2.5" cylinder during a 2 min. 15 sec. icing encounter in the BFG tunnel are shown in Fig. 11. In this case the temperature was -4 ° C, the velocity was 150 kts and the liquid water content was 1.5 gm/m^3. The IR camera was located 56 cm above and 84 cm upstream of the cylinder as shown in Fig. 3. As a result the cylinder was observed from above the horizontal and only the top of the cylinder was in view. In the images, the flow would be from below and a thermal profile corresponding to the mid-plane of the tunnel is shown on the right.

Prior to exposure to the icing cloud, in Fig. 11a the cylinder was at a uniform temperature of -4 ° C corresponding to the tunnel stagnation temperature. When the spray was turned on the stagnation region quickly warmed to 0 ° C indicating wet growth. The wet region spread for approximately 75 sec until it reached equilibrium covering the foreword 50° of the cylinder. The equilibrium wet region can be seen in Fig. 11b. As horned growth began to appear in the wet region, the boundary of the wet region became irregular as exhibited in Fig 11c. The enhanced heat transfer and the variation resulting from the rough surface and the horns can be seen in Fig. 11d which was taken 10 sec. after the spray was shut off. At this time a small wet region can still be observed in the stagnation area due to the low heat transfer in that region.

The extent of the wet surface region was studied using the IR video system for a variety of conditions. It was found that both the surface roughness and horns were generated within the wet region. However, rime feathers were always observed in the dry region. The wet region was always centered around the stagnation line and its extent varied as a function of icing conditions, particularly temperature. The angular extent of the wet region is plotted as a function of temperature in Fig 12.

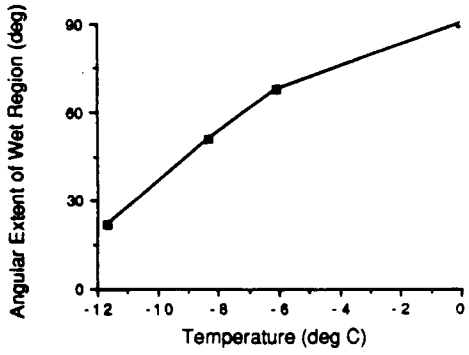


Fig. 12. Angular extent of wet region as a function of temperature for 2.5" cylinders in the BFG tunnel, (V=150 kts., W-1.0 g/m^3).

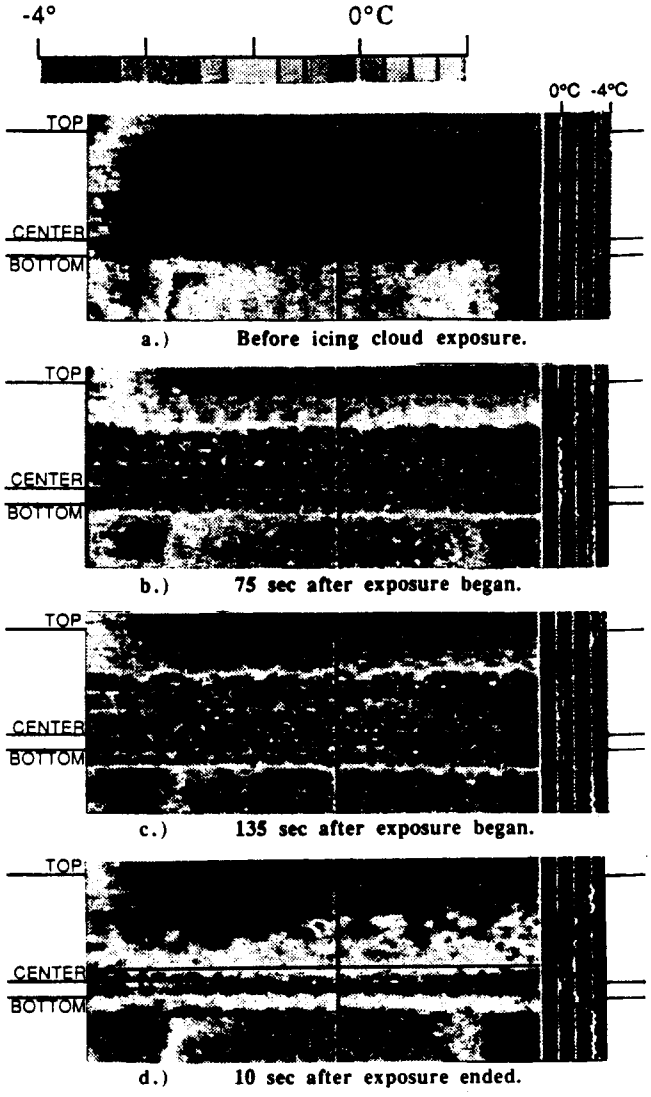


Fig. 11. Infrared thermographs of a 2.5" cylinder in the BFG tunnel, (T=-4°C, V=150 kts, W=1.5 g/m^3).

4.0 Preliminary Implementation of a Multi-Zone Model

4.1 Model Description

Based on the observations of surface water behavior during glaze ice accretion, a relatively simple modification to the existing ice accretion model is proposed which may improve the current model's accuracy within the glaze ice regime. In this proposed "Multi-Zone" model, the accreting ice surface is divided into two or more discrete zones which have varying surface water behavior and surface roughness. This is in contrast to current techniques which assume that the surface has uniform roughness and surface water runback.

In the simplest version of the model, the surface is divided into two zones. There is a "smooth" wet zone corresponding to the smooth zone described in Section 3.1. It is centered about the stagnation region and thin film runback occurs within the zone. In this region, the heat transfer would be that for a smooth surface and the Messinger[7] type runback model used in existing codes such as LEWICE appears to be valid. No change in the existing codes is considered necessary in the smooth zone.

The remaining ice surface would consist of a "rough" zone where surface tension effects are important and characteristic water beads or roughness elements appear. Within this rough zone the existing Messinger runback model is not appropriate for two reasons. First, the increased roughness will enhance the heat transfer and consequently increase the freezing fraction. Second, experimental observations by Olsen and Hansman[3,4,8,9] indicate that in some cases, surface water does not run back in the rough zone. In most cases surface tension forces dominate the water behavior and the beads remain stationary on the surface. At warm temperatures, high liquid water contents or high velocities, the beads may grow large enough that aerodynamic forces cause the beads to either slide aft on the surface or be torn from the surface and shed into the airflow.

In the absence of runback, the freezing fraction can be assumed to be unity within the rough region. This assumption obviates the need to calculate the convective heat transfer coefficient and the effective surface roughness which greatly simplifies the ice accretion model. In the presence of runback or sliding, the freezing fraction is less than unity and must be computed by the model through the Messinger heat balance approach with appropriate heat transfer coefficients corresponding to the enhanced roughness. In some cases it may be adequate to approximate the spatial variation of the freezing fraction by linear interpolation between the smooth zone value and unity at the point where rough zone ends or the surface is dry.

In the presence of shedding, the modeling is more complex. Some mechanism must be included to model the shedding process and the increased roughness within the rough zone. One approach would be to determine a maximum unfrozen water thickness for each element based on the local flow velocity. If the water layer in the element exceeds this value the water is shed from the surface in the model to both the runback and heat transfer. The roughness within the rough zone could then be modeled as the maximum stable droplet size.

Because the boundary between the rough and smooth zones moves with time it is necessary to incorporate this into the model . Since most ice accretion codes run a limited number of time steps, it was found to be convenient to include a third transition zone between the rough and smooth zones as shown in Fig.13. For cases where there is no runback in the rough zone the transition extends between the location of the rough to smooth transition at the beginning and end of the time step. For cases where there is runback in the rough zone the transition region would extend between the rough to smooth zone transition to the end of the rough zone or the point at which there is no further runback. Under these assumptions, the freezing fraction can simply be ramped up in the transition zone from the smooth zone value to unity. It should be noted that, in principle, the surface roughness could be varied instead of the freezing fraction. This was attempted; however, difficulties were encountered in correlating the effective heat transfer and equivalent sand grain roughness used in LEWICE with physical values. Therefore, the freezing fraction approach was employed in these preliminary evaluations.

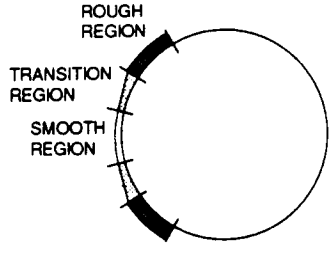
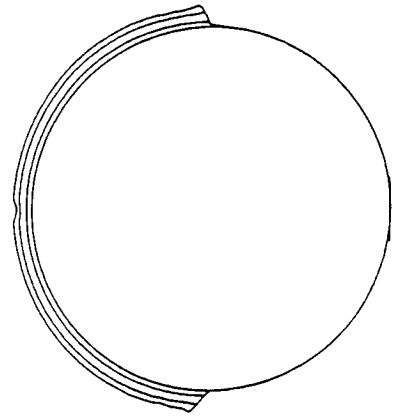


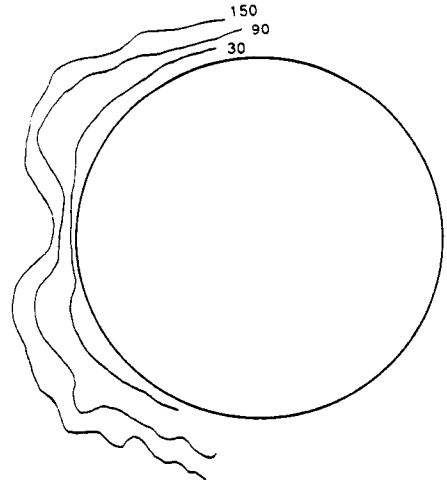
Fig. 13. Schematic representation of zones used in the model.

4.2 Results

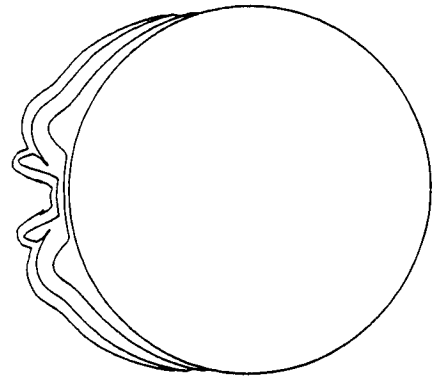
A prototype of the proposed "multi zone" model was implemented in the LEWICE code to evaluate the effectiveness of the concept. In this preliminary model the rough to smooth transition location and the extent of the wet growth region were taken from experimental observations. The intermediate and



a.) **Normal LEWICE at 45, 105, and 150 secs.**



b.) **Experimental result at 30, 90, and 150 secs.**



c.) **Modified LEWICE at 45, 105, and 150 secs.**

Fig. 14. Comparison of predicted and experimental ice shapes for a 1" cylinder, (T=-4.5°C, V=150 kts, W=0.35 g/m^3).

final ice accretions profiles were compared with experimentally measured profiles for several cases.

The first comparison was conducted for a low liquid water case of 0.35g/m^3 on a 1"cylinder in the DP tunnel at -4.5 ° C, and 150 kts Since there was little or no runback in this case the freezing fraction was ramped linearly through the transition zone from the smooth zone value to unity in the rough zone. The comparisons between the experimentally measured ice shapes and those predicted by the multi-zone LEWICE code and the normal LEWICE code with the Ruff roughness correlations [10] are shown in Fig. 14. It can be seen that there is a significant improvement in fidelity of the overall ice shape for the multi-zone case. There is however an irregularity in the ice surface which appears in the second time step and is thought to be an artifact of the ramped freezing fraction combined with the long time steps. The irregularity is located at the point where the freezing fraction becomes unity. Because of the relatively long time steps used in the simulation runback water collects at this point resulting in the observed bulge. Efforts are under way to minimize this effect in a physically realistic and numerically acceptable manner. In general, however the multi-zone LEWICE predicts the essential components of the ice accretion whereas the normal LEWICE prediction is not at all representative of the experimentally observed ice shape.

The second comparison was conducted for a high liquid water case of 1.5g/m^3 on a 2.5" cylinder in the BFG tunnel at 4.0 ° C, and 150 kts. In this case the rough to smooth transition reached the stagnation point by 60 sec as shown in Fig. 15. After this time the cylinder was uniformly rough with a roughness element size estimated at 2 to 3 mm from the grazing angle video. Runback was observed in the rough zone and the wet region was observed by the IR video system to extend to 60°. Since there was runback in the rough zone, the transition region in the multi-zone model extended from the rough to smooth transition point to the observed dry surface at 60°. It should also be noted that the liquid water content was reduced by a factor of 2 during the first 30 second time step to account for start up transients in the spray system.

The comparisons are shown in Fig. 16. There is excellent agreement between the multi-zone model and the experimental ice shapes for the first 60 sec while there were multiple roughness zones on the icing surface. The normal LEWICE code significantly underpredicted the ice accretion. After 60 sec simulation of the further accretion was simulated by running the normal LEWICE code on top of the 60 sec ice shape using an equivalent sand grain roughness of 3 mm corresponding to the observed physical roughness. This resulted in an extremely low freezing fraction of approximately 0.005 over the entire surface and generated almost no further accretion. Larger values of surface roughness were also used; however,the freezing fraction did not increase significantly.

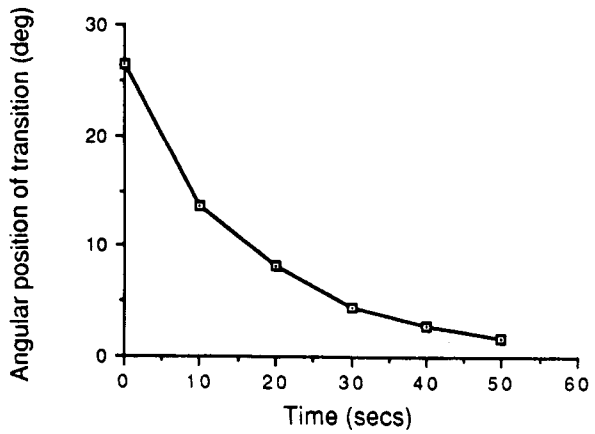
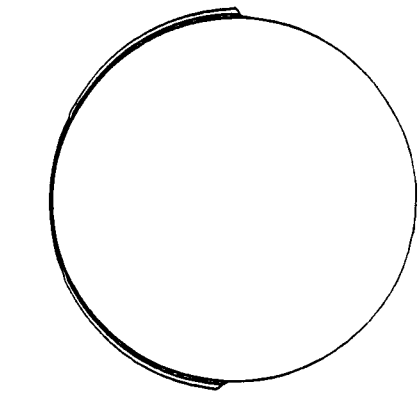
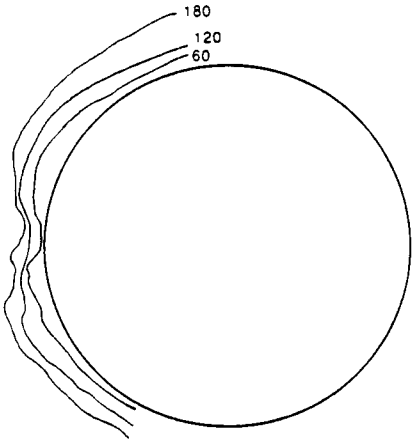


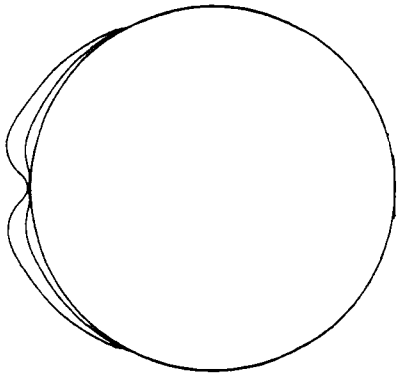
Fig. 15. Angular position versus time of smooth-rough transition location for the high liquid water content example.



a.) Normal LEWICE at 30 and 60 secs.



b.) Experimental result at 60, 120, and 180 secs.



c.) Modified LEWICE at 30 and 60 secs.

Fig. 16. Comparison of experimental and predicted ice shapes for a 2.5" cylinder, (T=-4°C, V=150 kts, W=1.5 g/m^3).

The LEWICE code clearly underpredicted the heat transfer for this case. In the normal LEWICE code surface roughness is used to adjust the heat transfer model. This illustrates the need for an improved treatment of the heat transfer if surface roughness is to be deterministically included in ice accretion models.

5.0 Conclusions

The investigation into modeling of surface roughness effects on glaze ice accretion has resulted in the following:

1. Accreting glaze ice surfaces have been observed to have distinct zones of surface roughness. They include: a smooth wet zone in the stagnation region where uniform film runback occurs, a rough zone where surface tension causes coalescence of the surface water into beads, a horn zone which can emerge out of the rough zone with large roughness elements, a runback zone where surface water runs back as rivulets, and a zone where rime feathers grow.

2. The location of the transition point between the smooth and rough zone was observed to propagate with time towards the stagnation point.

3. The initial location of the surface roughness transition point appears to be related to the transition of the boundary layer from laminar to turbulent. The propagation of the transition point appears to be related to the rate of bead formation at the interface between the rough and smooth zones.

4. Infrared video observations were made of accreting ice surfaces. This technique was used to investigate the extent of wet regions on the ice surface and also measured heat transfer enhancements resulting from roughness and horn elements.

5. Based on the above, a preliminary modification to the LEWICE ice accretion code which incorporates multiple roughness zones was proposed.

6. The proposed multi-zone modifications were implemented using experimental data to determine factors such as the rough to smooth transition locations. The multi-zone modifications significantly improved the performance of the code as compared with the normal LEWICE code. Based on the good performance of the preliminary multi-zone code, further efforts to develop a more fully deterministic and operational version of the code appear to be warranted.

Acknowledgments

This work was supported in part by the National Aeronautics and Space Administration and the Federal Aviation Administration under Grants NAG-3-666 and NGL-22-009-640. The work was also supported by the National Science Foundation Presidential Young Investigators Program, Award No. 8552702. Use of the icing wind tunnel facilities were provided courtesy of Data Products of New England, B. F. Goodrich Deicing Systems and the NASA Lewis Research Center. Dr. James Riley of the FAA Technical Center provided valuable input to the modeling efforts.

References

1. Achenbach, E., "The Effect of Surface Roughness on the Heat Transfer from a Circular Cylinder to the Cross Flow of Air," *International Journal of Heat Mass Transfer*, Vol. 20, 1977, pp. 359-369.

2. MacArthur, C.D., "Numerical Simulation of Airfoil Ice Accretion," AIAA Paper 83-0112, Jan. 1983.

3. Gent, R.W. and Cansdale, J.T., "The Development of Mathematical Modeling Techniques for Helicopter Rotor Icing," AIAA Paper 85-0336, Jan. 1985.

4. Olsen, W., and Walker, E., "Closeup Motion Pictures of the Icing Process," NASA LeRC Film, 1983.

5. Olsen, W.A., and Walker, E.,"Experimental Evidence for Modifying the Current Physical Model for Ice Accretion on Aircraft Structures", NASA TM 87184, 1987.

6. von Doenhoff, A.E., and Horton, E.A., "A Low Speed Experimental Investigation of the Effect of Sandpaper Type of Roughness on Boundary-Layer Transition", NACA TN 3858, 1956.

7. Messinger, B.L., "Equilibrium Temperature of an Unheated Icing Surface as a Function of Airspeed," *Journal of the Aeronautical Sciences*, Jan. 1953, pp. 24-42.

8. Hansman, R.J., and Turnock, S., "Investigation of Surface Water Behavior During Glaze Ice Accretion", AIAA Paper 88-0115, Jan. 1988.

9. Hansman, R.J., and Turnock, S., "Investigation of Microphysical Factors Which Influence Surface Roughness During Glaze Ice Accretion", *Proc. 4th Intl. Workshop on the Atmospheric Icing of Structures*, Sept. 1988.

10. LEWICE Users Manual, 1988.

ULTRASONIC TECHNIQUES FOR AIRCRAFT ICE ACCRETION MEASUREMENT * ^

R. John Hansman, Jr.,* and Mark S. Kirby†
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, Massachusetts

and

Fred Lichtenfelts ‡
Simmonds Precision
Aircraft Systems Division
Vergennes, Vermont

ABSTRACT

Results of tests to measure ice growth in natural (flight) and artificial (icing wind tunnel) icing conditions are presented. Ice thickness is measured using an ultrasonic pulse-echo technique. Two icing regimes, wet and dry ice growth, are identified and the unique ultrasonic signal characteristics associated with these different types of ice growth are described.

Ultrasonic measurements of ice growth on cylinders and airfoils exposed to artificial and natural icing conditions are presented. An accuracy of ±0.5 mm is achieved for ice thickness measurement using the pulse-echo technique. The performance of two-probe type ice detectors is compared to the surface mounted ultrasonic system. The ultrasonically measured ice accretion rates and ice surface condition (wet or dry) are used to compare the heat transfer characteristics for flight and icing wind tunnel environments. In general the heat transfer coefficient is inferred to be higher in the wind tunnel environment, not likely due to higher freestream turbulence levels. Finally, preliminary results of tests to measure ice growth on airfoil using an array of ultrasonic transducers are described. Ice profiles obtained during flight in natural icing conditions are shown and compared with mechanical and stereo image measurements.

1.0 INTRODUCTION

The direct measurement of ice thickness on critical aircraft surfaces, in flight and on the ground, has proven to be a difficult task. Current ice detectors generally measure the accretion of ice on a probe and infer the accretion on other aircraft components. In many applications this results in significant uncertainty as to the actual ice accretion on critical components. Recently, ultrasonic techniques have been developed which allow the ice accreted on critical components to be measured directly by a transducer imbedded in the surface. This technology has significant implications for operational use in measuring ice thickness and accretion rate in flight and automatically controlling ice protection systems. The technique also has potential applications in detecting hazardous ground icing events. In addition, ultrasonic techniques can be used to aid in the certification of aircraft for flight in icing condition and in understanding the physics which underlie the ice accretion problem.

The purpose of this paper is to present an overview of recent work on the measurement of ice thickness using ultrasonic pulse-echo techniques. Additional ultrasonic methods have been investigated including ice bond monitoring by shear wave techniques and thin ice detection by pulse-echo amplitude measurement. In the interest of brevity only the pulse-echo results will be reviewed here. Experimental measurements of ice growth in icing wind tunnel and flight icing conditions are presented. The ultra-sonic pulse echo technique is shown to provide the potential for the development of an operational instrument. In addition, the technique provides previously unobtainable resolution on the temporal and spatial growth of ice accretion in simulated and natural icing conditions. Finally, the unique capability of the ultrasonic pulse-echo technique to detect the presence of liquid water on an accreting ice surface allows the threshold between wet and dry ice growth to be experimentally determined. It is shown that this wet/dry threshold data may be used to infer valuable information about the heat transfer from the ice surface, and thus aid in understanding the ice accretion process.

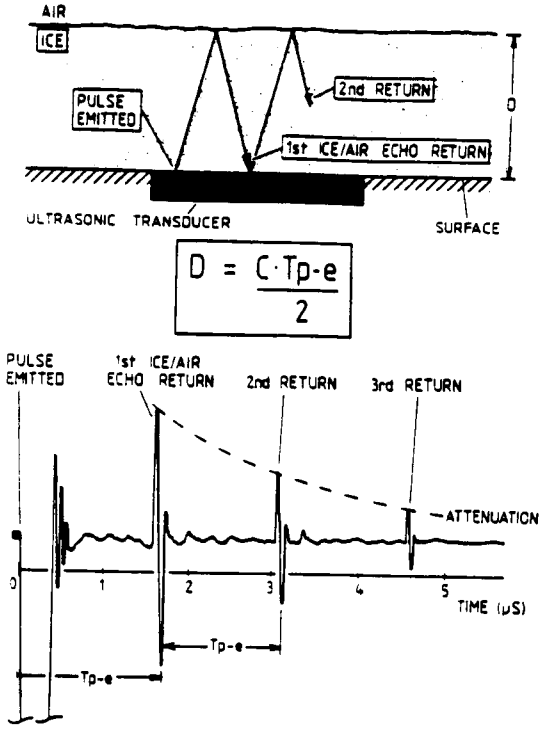


Fig. 1 Ultrasonic pulse-echo thickness measurement and typical ultrasonic pulse-echo signal in ice.

---

* Associate Professor, Member AIAA.
† Research Assistant, Currently at Rolls Royce, Bristol England.
‡ Lead Systems Engineer, Member AIAA.

*AIAA-84-4656-CP.
^Some original figures were not available at time of publication.

2.0 ULTRASONIC PULSE-ECHO MEASUREMENT TECHNIQUES.

2.1 ULTRASONIC PULSE-ECHO THICKNESS MEASUREMENT

Ultrasonic pulse-echo measurement of ice thickness on a surface is accomplished by emitting a brief compression or shear wave from a small ultrasonic transducer mounted flush with the accreting surface (Fig. 1). The pulse travels through the ice, is reflected at the ice/air interface and then returns to the transducer as an echo signal. The time elapsed, $T_{p\text{-}e}$, between the emission of the pulse from the transducer and the return of the echo from the ice interface can then be used to calculate the ice thickness, D, from the formula:

$$D = \frac{C \cdot T_{p\text{-}e}}{2} \qquad (1)$$

where C is the speed of propagation of the pulse-echo signal in ice. In a previous study[1], the speed of propagation was found to be insensitive to different types of ice (glaze, rime and mixed) formed at typical flight airspeeds. Fig. 2 illustrates the approximately constant speed of sound observed experimentally in ice samples formed under different icing conditions. A value of 3.8 mm/μs was used for the speed of sound in ice for all the results presented in this paper.

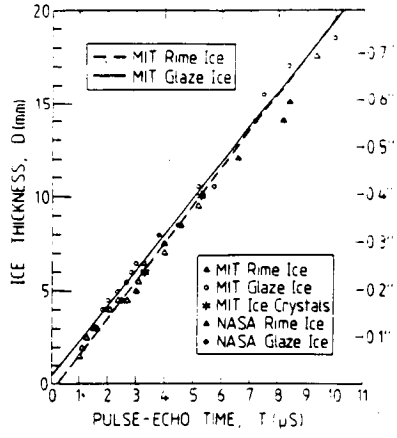


Fig. 2 Ice thickness vs. pulse-echo time constant speed of sound for ice samples formed under different icing conditions.

2.2 ULTRASONIC SIGNAL CHARACTERISTICS FOR DRY ICE GROWTH

In cold conditions, all the droplets impinging at a given location on a body freeze on impact and the ice surface formed are "dry". During dry ice growth, the ice surface tends to remain relatively smooth and simply increases in thickness. The received echo, therefore, appears to translate in time with a velocity proportional to the accretion rate. This behavior is illustrated in Fig. 3. Since the surface formed by dry ice growth is typically smooth, a sharp, well-defined echo is received, as shown in the figure.

2.3 ULTRASONIC SIGNAL CHARACTERISTICS FOR WET ICE GROWTH

In warm conditions at temperatures just below 0° C, the heat transfer from the accreting surface is insufficient to completely freeze all the impinging droplets and liquid water will be present on the ice surface. This form of ice accretion will be referred to as "wet ice growth", and is characteristic of glaze and mixed ice formation.

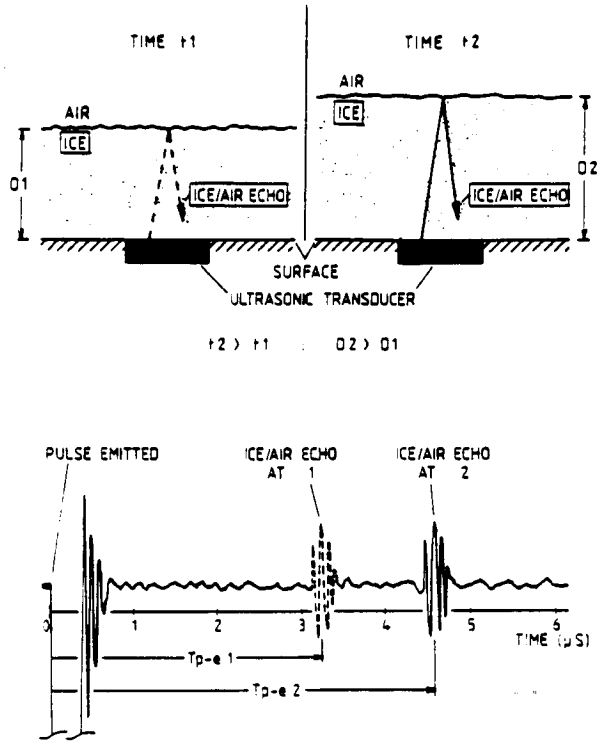


Fig. 3 Ultrasonic signal characteristics for dry ice growth.

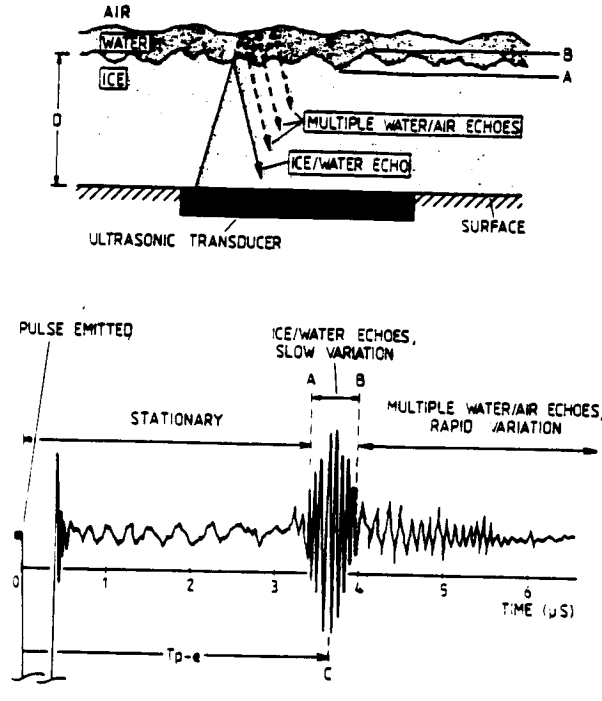


Fig. 4 Ultrasonic signal characteristics for wet ice growth.

During wet ice growth the ice surface is covered, at least partially, by a thin liquid layer. The ultrasonic pulse thus encounters two separate interfaces: an ice/water interface and a water/air interface, as shown in Fig. 4. Two echo signals are therefore received by the transducer; the first from the ice/water interface and a second, later echo from the water/air interface. The echo received from the ice/water interface is characteristically broader than the echoes received during rime ice growth due to the rougher ice surface formed during glaze icing. As in the rime ice case, this ice/water interface echo translates in time as liquid freezes at the interface and increases the accreted ice thickness.

Since the liquid layer over the ice surface is thin, and its attenuation is low, multiple echoes are received from the water/air interface following the ice/water echo. Due to distortion of the liquid surface by impinging droplets and the external flow, the detailed shape of these multiple water/air echoes changes rapidly as the instantaneous local thickness and surface orientation of the liquid layer varies. This rapid variation in the received echo signal is absent when the ice surface is dry. Wet and dry ice growth may thus be distinguished by observing the time variation of the received ultrasonic echo signals.

3.0 MEASUREMENT OF ICE GROWTH DURING ARTIFICIAL AND NATURAL ICING CONDITIONS

This section presents results of tests to measure and compare ice growth in both icing wind tunnel and icing conditions. Ice thickness on the stagnation line of a cylinder was measured as a function of the exposure time, using the ultrasonic pulse-echo technique in the NASA Lewis Icing Research Tunnel and on the NASA Icing Research aircraft. Preliminary tests of digitally processed "real time" ultrasonic measurement were conducted in the Boeing Icing Tunnel and on the Boeing 757 test aircraft. In addition, the ultrasonic signal characteristics were used to distinguish the presence of liquid water on the accreting ice surface. These wet/dry ice growth observations are then used to compare the heat transfer processes occurring in the icing wind tunnel and in flight.

3.1 EXPERIMENTAL APPARATUS

The experimental apparatus used for both the NASA Lewis Icing Research Tunnel (IRT) tests and the NASA natural icing flight tests consisted of the following equipment:

1. A cylinder containing small ultrasonic transducers mounted flush with the cylinder surface.

2. A pulser/receiver unit to drive transducers.

3. An oscilloscope to display the resulting pulse-echo signal.

4. A video camera/recorder to record the pulse-echo signal displayed on the oscilloscope.

Fig. 5 schematically illustrates the configuration of the experimental apparatus. For the IRT tests a 10.2 cm diameter cylinder was instrumented with four ultrasonic transducers. The transducers used had plane circular faces with element diameters of either 0.6 or 0.3 cm. The transducers had center frequencies of 1, 2.25 and 5 MHz; all were broadband, heavily damped transducers. As expected, the 5 MHz transducers provided the highest thickness resolution (due to the shorter wavelength), and all results presented are for 5 MHz transducers. The transducers were placed on the stagnation line and were located near the midsection of the cylinder. A pulser/receiver unit provided the electrical signals necessary to generate the ultrasonic pulse and amplify the return echo.

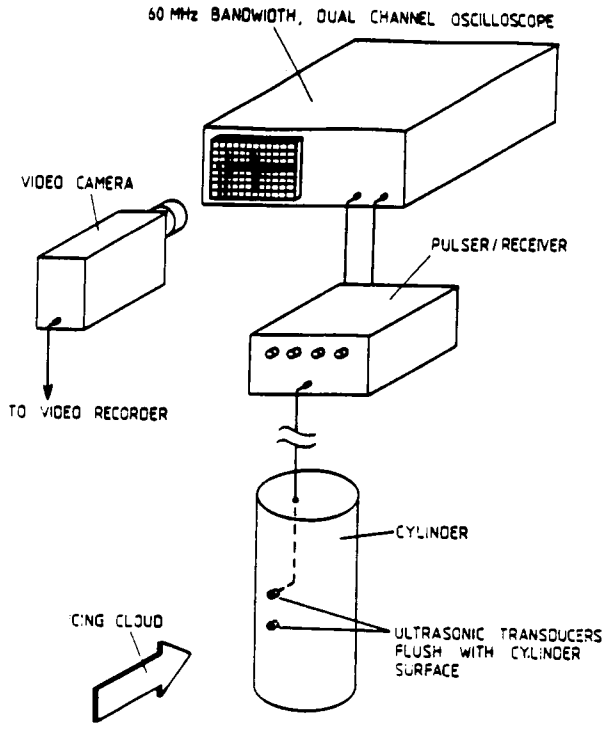


Fig. 5 Schematic of experimental apparatus configuration.

The cylinder employed for the NASA natural icing flight tests was an 11.4 cm diameter cylinder. This cylinder was instrumented with two 0.6 cm diameter, 5 MHz transducers, mounted in the stagnation line of the cylinder. Again a single pulser/receiver unit was used to drive either transducer.

In order to observe the ultrasonic signal characteristics in detail, an oscilloscope was used in both series of NASA tests to display the pulse-echo signals. A video camera was used to permanently record the oscilloscope signal for subsequent analysis. The video camera simultaneously recorded the cylinder exposure time from an electronic clock.

Preliminary tests of a digitally processed "real time" ultrasonic measurement techniques were conducted on a full scale model of a 757 leading edge slat in the Boeing Icing Tunnel. The slat was instrumented with commercially available 5 MHz transducers. Single and dual element transducers with polystyrene delay lines were tested. The diameter of the exposed polystyrene face was 0.63 cm. Transducer excitation and return echo signal conditioning was provided by commercial ultrasonic pulser/receiver hardware. A digital oscilloscope was used to capture return echoes and measure the round-trip delay. A micro-computer subsequently computed the ice thickness (approximately every 2.5 seconds) and logged the data on a printer. In addition, the oscilloscope pulse-echo waveforms were also video recorded to allow visual validation of the computer processing.

The Boeing natural icing flight tests used custom 2.25 MHz transducers with an exposed diameter of approximately 1.3 cm. One transducer was mounted on an outboard leading edge slat and another on a 7.6 cm diameter fared cylindrical shaped mast extended approximately 40 cm from the forwardmost passenger window as shown in Fig. 6. The transducers were connected to a modified commercial ultrasonic thickness gage which provided a digital display as well as an analog thickness output. In addition, echoes were monitored on an oscilloscope. Strip chart recordings were

made of the aircraft altitude, true airspeed, outside air temperature (OAT), liquid water content (LWC) and the analog ice thickness output. Visual observations of a wing mounted "barber-pole" were made to confirm the ice thickness readings.

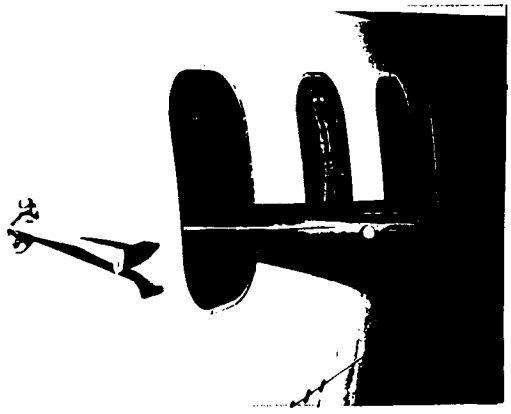


Fig. 6 Installation of the ultrasonic detector in the 757 window.

3.1.1 Icing Research Tunnel Tests

The instrumented cylinder was vertically suspended from the roof of the NASA Icing Research Tunnel (IRT), as shown schematically in Fig.7. The cylinder was exposed to the icing cloud, and the pulse-echo signals from the transducers being compared recorded using the video-camera/recorder. Each exposure was made with constant icing cloud parameters and typically lasted eight minutes. At the completion of a run, the iced diameter of the cylinder at each transducer located was measured mechanically using a pair of outside calipers. The ice thickness over the transducers was inferred from this measurement for comparison with the ultrasonically measured ice thickness. The cylinder was then completely de-iced in preparation for the next run. Ice growth for a total of fifteen different icing cloud parameter sets, ranging from glaze to rime icing conditions, was recorded using the ultrasonic pulse-echo system.

3.1.2 Boeing Icing Tunnel Tests

The 757 leading edge model spanned the tunnel at approximately a 20° sweep angle. A total of 20 tests were conducted. The test conditions covered a temperature range of -28° C to +2° C; LWC of 0.2 to 0.9 gm/m^3 and airspeed from 100 to 175 MPH. Droplet size (MVD) was gravitational 10 or 20 microns, depending upon LWC. The exposure time varied from two to 15 minutes. Typically ice was allowed to accrete until between 0.150 to 0.300 inches were present. The tunnel was then opened and a manual measurement of the ice thickness over the transducer was made using calipers.

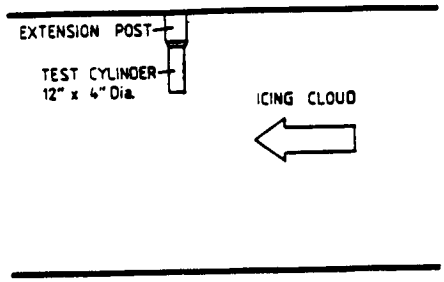


Fig. 7 Test cylinder installation for icing tunnel (IRT) tests.

3.1.3 Natural Icing Tests

Natural icing tests were conducted in flight from the NASA Lewis Icing Research Aircraft, a De Havilland Twin Otter. The 11.4 cm cylinder instrumented with ultrasonic transducers was exposed to icing conditions at the end of an extension post that was vertically extended through the roof of the aircraft by an experiment carrier mounted in the aircraft (see Fig.8). The cylinder midsection was located 53.3 cm into the freestream when extended. The cylinder was typically exposed throughout the entire icing encounter, the pulse echo signal again being continuously displayed and recorded by the oscilloscope/camera system. At the end of the encounter the cylinder was retracted into the aircraft and the ice thickness over the transducer measured using outside calipers. Four research flights were conducted with the ultrasonic system during the period March 30 to April 2, 1985.

The Boeing 757 flight tests were conducted on February 28, 1987 from Boeing Field in Seattle. Four icing encounters occurred at an altitude of approximately 10,000 ft. Each encounter lasted from five to nine minutes and was terminated by climbing above the icing conditions. Following the last encounter, total ice accumulation was in excess of one inch. The transducer mounted on the mast accurately tracked ice accretions in excess of 1 inch. The slat mounted transducer failed to perform as expected. This is thought to be due to an inadequate transducer mounting arrangement which prevented the direct bonding of accreted ice to the sensor.

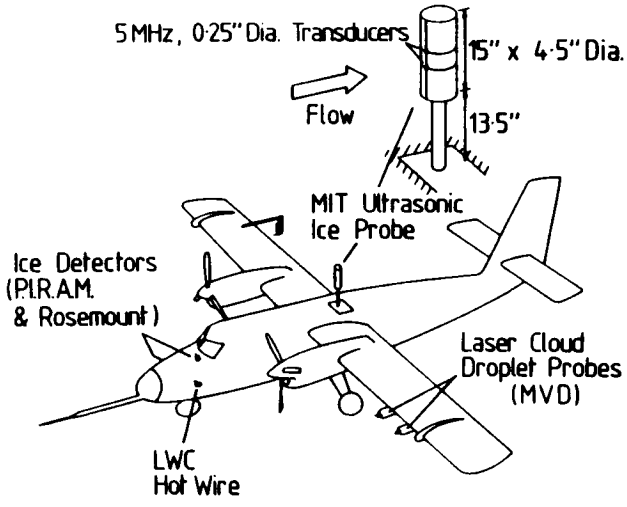


Fig. 8 Test cylinder installation for natural icing (NASA Lewis Twin Otter aircraft) tests.

3.2 RESULTS AND DISCUSSION

3.2.1 Icing Research Tunnel Tests

Figs. 9 and 10 show ice thickness measured with the ultrasonic system plotted against exposure time for two different icing cloud parameter sets, labeled "heavy" and "light" icing respectively. The heavy icing conditions represent an icing cloud with a high liquid water content of approximately 0.77 gm/m^3 and a large median volume droplet size of 20 microns; whereas the light icing conditions represent relatively lower values for both of these parameters. The freestream airspeed was 102.8 m/sec for both cases shown. Ice growth at three different temperatures (-8.0° C, -17.5° C and -28.6° C), is shown for both icing conditions.

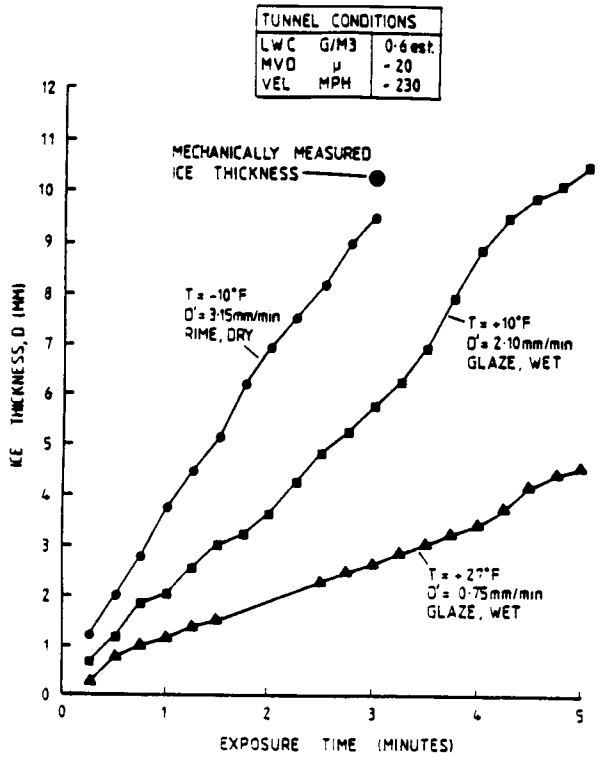


Fig. 9 Ice growth under "heavy" icing tunnel conditions.

The total ice accretion was mechanically measured with outside calipers at the completion of each run and is also plotted, if known. The accuracy of the ultrasonic pulse-echo thickness measurements was found to be within 0.5 mm of the accretion measured with the calipers for all runs where mechanical measurements were obtained. This good agreement between the mechanically measured thicknesses and the ultrasonic measurements confirmed previous experimental results which showed the speed of propagation of the ultrasonic signal to be insensitive to widely varying icing conditions.

The ice accretion rates measured in the IRT (given by the slope of the thickness versus time curve) remained fairly constant throughout each run. This behavior is a result of both the constant icing cloud conditions in the tunnel and the location of the transducer on the cylinder stagnation line

While observed accretion rates are relatively constant throughout each run, the accretion rates differ significantly for the different icing cloud conditions. Due to the higher impinging mass flux, the accretion rates measured under the "heavy" icing conditions were greater than those observed under the "light" icing conditions at the same temperature. For example at -28.6° C the accretion rate is 3.15 mm/min for the heavy icing conditions while it was only 1.05 mm/min for the "light" icing case. Of more interest is the effect of the cloud temperature on the stagnation accretion rate. At the relatively warm temperatures just below 0° C, the accretion rates for both the heavy and light cases are low, and the ice growth is observes to be wet. Under these conditions, the rate of heat removal from the ice surface was sufficient to completely freeze all the locally impinging droplets, liquid forms on the accreting surface and ran back away from the stagnation region. However, as the temperature of the icing cloud was progressively reduced, the accretion rate was observed to increase to a maximum value and then remain constant despite further decreases in icing cloud temperature.

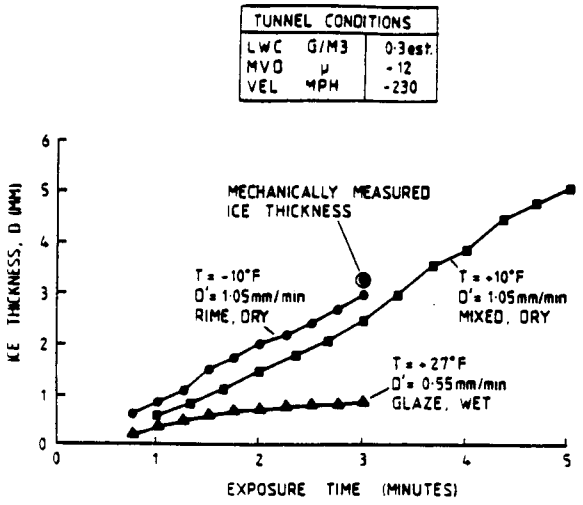


Fig. 10 Ice growth under "light" icing tunnel conditions.

The lower accretion rates observed at the warmer cloud temperatures are due to liquid runback from the stagnation region. As the cloud temperature is reduced, the rate of heat removed from the ice surface increases, this results in the freezing fraction (ratio of impinging to freezing water flux) and the accretion rate increasing. For dry growth, the freezing fraction is unity and the accretion rate is a maximum (for the particular cloud parameters). This behavior is illustrated in Fig. 11, which shows the average accretion rates measured for the heavy and light icing conditions as a function of the cloud temperature, T_∞. For the light icing case the accretion rate increases with decreasing temperature to -17.5° C, at which point dry ice growth is observed, the accretion rate then remains constant as the icing cloud temperature is further decreased to -28.6° C. However, for the heavy icing conditions it can be seen that at -17.5° C the growth is still wet and the accretion rate continues to increase from -17.5° C to -28.6° C. For these "heavy icing" conditions the heat transfer from the ice surface as still insufficient at -17.5°C to completely freeze the high impinging mass flux, and hence wet growth was observed. At -28.6° C though the ice growth is dry and the accretion rate measured, 3.15 mm/min is therefore the maximum value for those conditions.

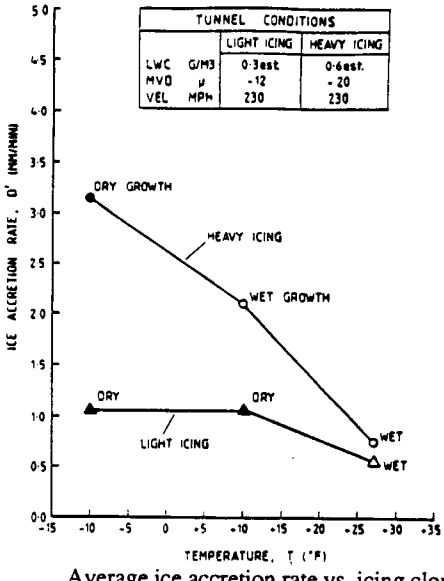


Fig. 11 Average ice accretion rate vs. icing cloud temperature for "heavy" and "light" icing tunnel conditions.

3.2.2 Boeing Icing Tunnel Tests

Fig. 12 shows an example of ice thickness measured and processed in real time with the digitally analyzed ultrasonic data. As can be seen in the data, the preliminary algorithms used in this test required approximately 0.075" of ice before it locked in. However, a clear reaction in the echo amplitude was always observed within 5 seconds of spray initiation. This amplitude behavior has been subsequently used for ultrasonic measurement of thin ice accretions. The agreement between final ultrasonic and mechanical ice thickness measurements was generally good with most cases agreeing to better than 10%. Those outliers which did occur were trackable to problems with the preliminary signal processing algorithm.

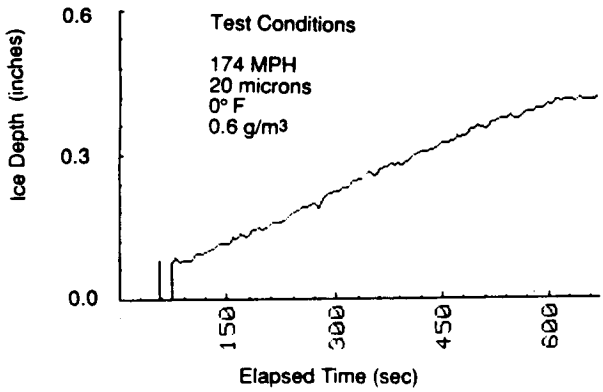


Fig. 12 Digitally processed ice growth measurement.

3.2.3 Natural Icing Tests

Fig. 13 contains a summary of the time averaged icing conditions during each of the four in-flight exposures of the ultrasonic system obtained on the NASA Icing Research Aircraft. The shape of the accreted ice at the completion of each exposure was obtained from photographs of the iced cylinder. Also shown are the final ice thicknesses measured with the ultrasonic system and the outside calipers. The accuracy of the ultrasonic measurements is again seen to be within ±0.5 mm of the mechanically measured ice thickness.

Fig. 14 is a plot of the ice thickness measured with the ultrasonic system versus the cylinder exposure time, for research flight 85-24. Ice pulse-echo data, and the average accretion rate for the encounter was 0.88 mm/min.

Fig. 15 is a plot of icing cloud and ice accretion data measured by existing instrumentation on the aircraft, also for flight 85-24. The upper plot shows "icing rate" in mm/min measured by Rosemount and PIRAM ice detectors. Also shown is the ice accretion rate measured on the stagnation line of the test cylinder by the ultrasonic system. This accretion rate was determined by differentiating the measured ice accretion (Fig. 14) with respect to time. The lower plot shows the cloud liquid water content in gm/m^3 measured by a Johnson-Williams hot-wire probe. Due to the relatively cold cloud temperature and low liquid water contents observed during this encounter, the icing rate is expected to be approximately proportional to the cloud liquid water content.

From Fig. 15, it can be seen that three icing rate plots (Rosemount, PIRAM and Ultrasonic) show overall similarity to the LWC plot, (e.g., all contain a decrease in icing rate after two minutes, corresponding to a decrease in the cloud liquid water content). However, the magnitudes of the icing rates indicated by the three instruments are markedly different, the PIRAM and Rosemount respectively indicating average icing rates of 2.45 and 1.96 mm/min for the encounter while the

TIME - AVERAGED ICING CONDITIONS						
FLIGHT NUMBER	EXPOSURE TIME (MIN:SEC)	TEMP (°F)	LWC (GM/M3)	MVD (μ)	VEL (MPH)	ALTITUDE (FEET)
85-22	20:41	+22.3	0.36	11.9	157	2728
85-23	47.08	+26.1	0.18	11.4	147	3404
85-24	10.07	+5.7	0.46	14 est.	161	6480
85-25A	36:40	+4.8	0.19	14.1	167	7019
85-25B	11.29	+8.8	0.12	12.0	167	6837

FLIGHT NUMBER	ULTRASONICALLY MEASURED ICE THICKNESS (MM)	MECHANICALLY MEASURED ICE THICKNESS (MM)	ICE SHAPE
85-22	3.1	2.7	TRANSDUCER, FLOW, WET GROWTH
85-23	3.3	ICE SLID OFF	WET
85-24	8.9	8.7	DRY
85-25A	VIDEO TAPE DAMAGED	9.6	DRY
85-25B	1.0	1.3	DRY

Fig. 13 Summary of natural icing test results.

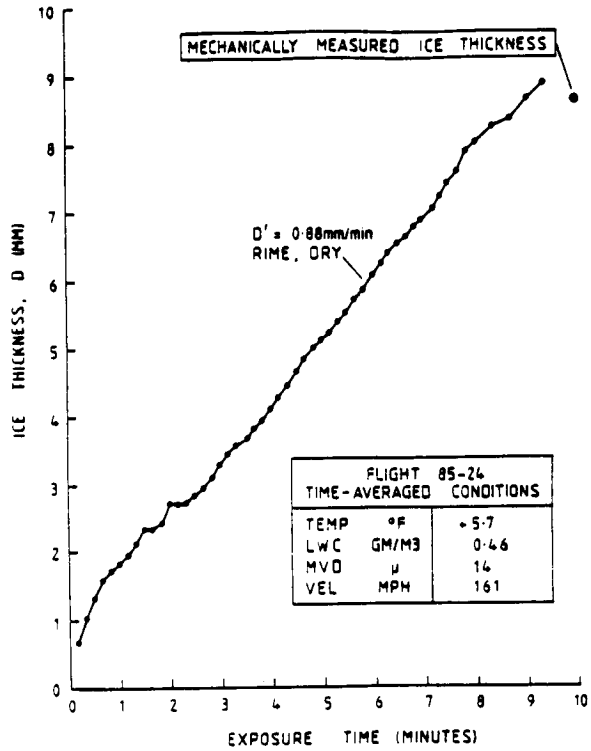


Fig. 14 Ice thickness vs. exposure time for flight 85-24.

average ultrasonically measured accretion rate is 0.88 mm/min. The ultrasonically measured icing rate is consistent with the mechanically measured accretion on the test cylinder of 8.7 mm at the end of the ten minute exposure. The higher icing rates recorded by the PIRAM and Rosemount systems appear to be inconsistent with the cloud liquid water content measured by the Johnson-Williams probe, a regression line analysis of measured icing rate versus Johnson-Williams cloud liquid water content was made which indicate that the PIRAM and Rosemount systems respectively imply icing rates of 2.06 and 1.36 mm/min at zero liquid water content.

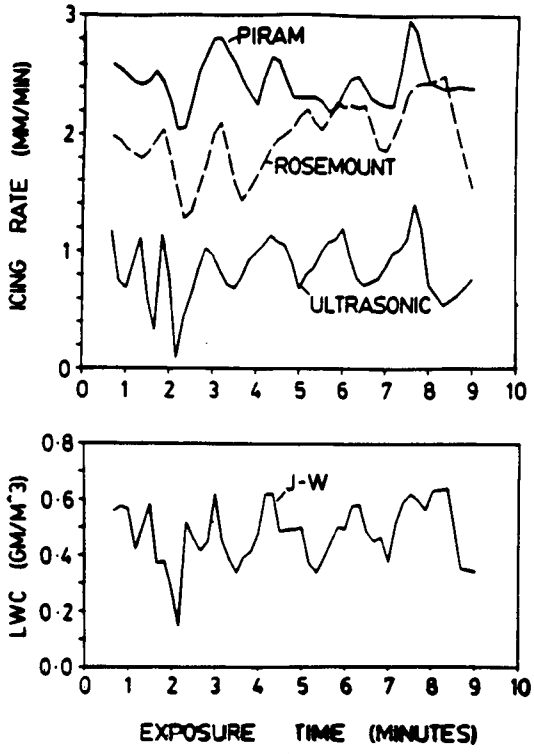


Fig. 15 Comparison of ice accretion rates measured by PIRAM, Rosemount and ultrasonic ice detectors and cloud liquid water content measured by a Johnson-Williams hot-wire probe for flight 85-24.

The PIRAM and Rosemount systems are both probe-type detectors that measure ice accumulation on a small, protruding probe. The calibration of these systems depends on the collection efficiency of the probe, which in turn varies with the cloud droplet size distribution. Since the ultrasonic system is non-invasive, ice thickness and icing rate are directly measured on the surface of interest and no calibration is required. Referring again to Fig. 15, it can be seen that the time variation of the liquid water content is recorded by the ultrasonic system while the Rosemount and PIRAM plots do not contain the same level of detail, particularly beyond four minutes exposure time, with the Rosemount data showing an apparent drift to a higher icing rate. The time response of probe-type systems is generally limited by two factors: the amount of ice necessary to initiate the icing signal and the need to repeatedly thermally de-ice the probe. The response time of both the PIRAM and Rosemount instruments is therefore dependent on the severity of the icing conditions and also physically limited by the time taken to de-ice the probe (typically 5-7 seconds). The time response of the ultrasonic system is not limited by the need for de-icing and does not depend on the icing severity; it is only fundamentally limited by the ultrasonic pulse repetition frequency, which is typically several KHz.

Fig. 16 presents strip chart data from an 8 minute period of the 757 flight test. The ultrasonically measured ice accretion is the bottom trace in Fig. 16. The measured ice accretion behavior is consistent with the measured liquid water content, also shown in Fig. 16. Reliable measurements were made to greater than 0.75 inches before the detector saturated. The average ambient temperature during the encounter was -15° C, the pressure altitude was 10,000 ft. and the true airspeed was 250 KTS.

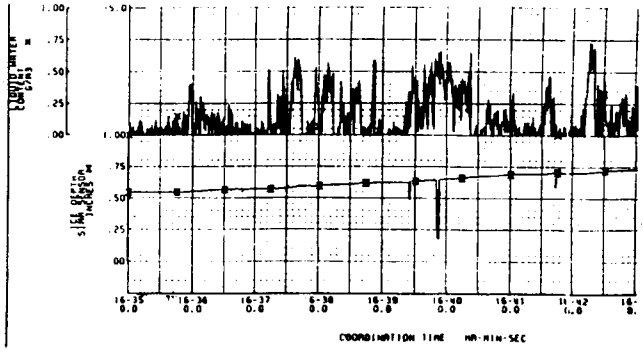


Fig. 16 Example of strip chart recordings of cloud liquid water content and ultrasonically measured ice thickness for the Boeing 757 tests.

3.3 HEAT TRANSFER MEASUREMENTS FOR AN ACCRETING ICE SURFACE

As discussed earlier, when the rate of heat removal from the icing surface is insufficient to freeze all the impinging droplets, liquid will form on the surface and the ice growth will be wet. By experimentally measuring the type of ice growth occurring (wet or dry) under particular icing cloud conditions, it is possible to parametrically infer limits on the heat transfer magnitude at the icing surface. A clear understanding of the heat transfer is essential if accurate analytic ice accretion models are to be developed. Due to experimental difficulties associated with heat transfer measurements on actual ice surfaces, local heat transfer coefficient measurements have only been made around wooden or foam models of typical ice shapes. In addition, experimental measurements comparing local heat transfer coefficients obtained in icing wind tunnels and in flight are essential if natural icing conditions are to be accurately simulated in icing wind tunnels. Very little experimental data exists in this area.

This section presents a comparison of experimentally measured wet and dry ice growth data with theoretical wet/dry threshold curves calculated using a steady-state energy balance for the stagnation region. The energy balance shown in Fig. 17 considers six modes of energy transport to and from the icing surface, the most significant terms being the heat added due to the latent heat of fusion and aerodynamic heating ("ram-rise"). The heat is removed primarily due to convection and evaporation or sublimation. The ultrasonically measured wet and dry ice growth data for the cylinders tested in the icing wind tunnel and in flight are used to compare a series of heat transfer coefficients for the cylinder stagnation region. Using this approach, the importance of freestream turbulence and surface roughness on the local heat transfer can be compared for artificial and natural icing conditions.

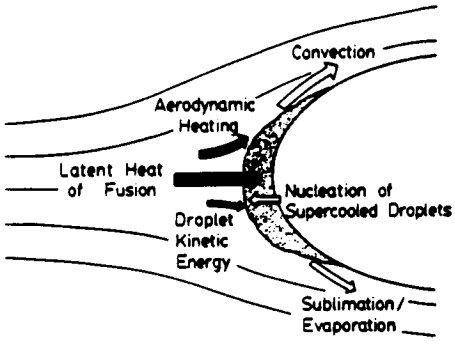


Fig. 17 Modes of energy transfer for an accreting ice surface.

The heat transfer coefficients used are from a recent experimental study by Van Fossen, et al[2], in support of the NASA icing research program. In the Van Fossen study, heat transfer coefficients were measured for two different freestream turbulence levels, 0.5% and 3.5%, and for two different cylinder surface conditions, one smooth and one roughened with grains of sand. A more complete discussion of these heat transfer models and the steady-state energy balance used to calculate the wet/dry threshold curves is contained in reference 3.

3.3.1 Icing Research Tunnel Results

Fig. 18 shows the ultrasonically measured ice growth for six different icing conditions in the icing research tunnel. The freestream velocity was 102.8 m/s (230 mph) for all six runs shown. Also shown are four wet/dry threshold curves calculated using the Van Fossen[3] heat transfer coefficients. These curves are plotted versus ambient temperature and were calculated for a freestream velocity of 102.8 m/s (230 mph), and a cylinder diameter of 0.102 m (4 in). The four curves shown thus represent the transition between wet and dry ice growth calculated for the four different local heat transfer coefficients implied by the Van Fossen data. If the local impinging liquid water content exceeds this critical value for a given ambient temperature then the ice growth is calculated to be wet, and if the impinging liquid water content is less than the critical value the ice growth is predicted to be dry. From the figure it can be seen that the heat transfer coefficient that best predicts the experimentally observed pattern of wet and dry ice growth is that measured for the cylinder roughened with sand and at a freestream turbulence level of 3.5%. While dry ice growth was observed at -28.6°C and an impinging liquid water content equal to 0.47 g/m^3, the heat transfer coefficients for the 0.5% freestream turbulence level clearly imply wet growth for these conditions. Thus it appears that the 0.5% turbulence level heat transfer coefficients underpredict the actual heat transfer in the icing tunnel, and are therefore too low.

Since the rough surface, 3.5% freestream turbulence heat transfer coefficient appeared to best approximate the actual heat transfer occurring in the icing research tunnel (based on the wet/dry ice surface data from th ultrasonic tests), this heat transfer model was compared with other ultrasonic wet/dry ice surface data obtained at different tunnel icing cloud conditions, with similarly favorable results.

Based on the results presented in Fig. 18, it appears that the heat transfer coefficient model that best approximates the actual heat transfer occurring in the icing research tunnel is the 3.5% turbulence level, rough surface model. The actual heat transfer coefficient is clearly greater than those applicable at the low (0.5%) turbulence level and may be even greater than that implied by the high (3.5%) turbulence level model.

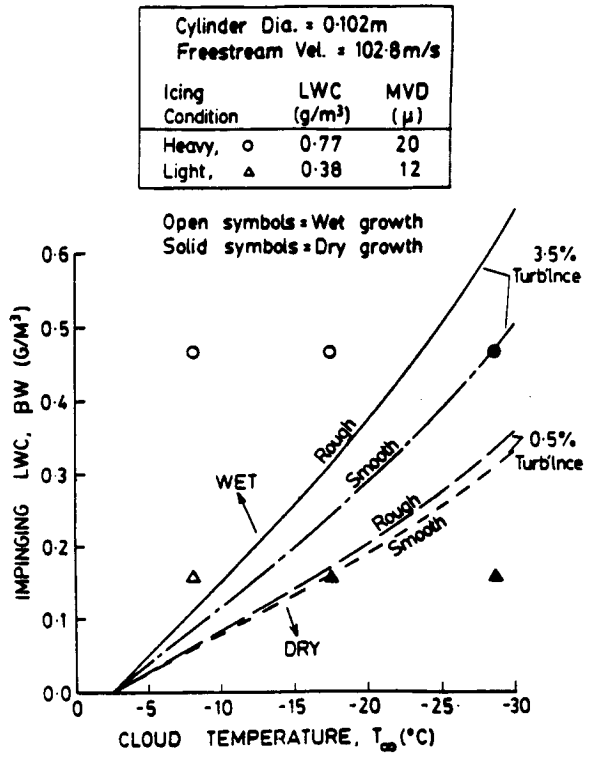


Fig. 18 Plot of impinging liquid water content vs. cloud temperature showing ultrasonically measured wet/dry ice growth and theoretical wet/dry threshold curves for four different heat transfer coefficients (V_∞= 102.8 m/sec.).

3.3.2 Natural Icing Cloud Test Results

While constant icing conditions were maintained throughout each exposure in the icing research tunnel, the natural icing cloud conditions, most noticeably the liquid water content, were not constant throughout each flight. Fig. 19 shows a plot of cloud liquid water content versus exposure time for research flight 85-24. The liquid water content was measured by a Johnson-Williams hot-wire probe located near the nose of the aircraft (see Fig. 8). Also shown, are the experimentally observed periods of dry, wet and transitional ice growth, produced by the varying liquid water content. The ultrasonic echo patterns received from the accreting ice surface were used to determine if the ice growth was wet, dry or transitional. Ice growth was characterized as transitional when the time variation of the ultrasonic echo pattern was between that characteristically observed for wet and completely dry ice growth.

Fig. 20 is a plot of impinging liquid water content versus cloud temperature. The experimentally observed ice growth regimes during the four research flights conducted are shown. Note that during flights 85-24 and 85-25 the full range of ice growth regimes were encountered with periods of dry, transitional and wet ice growth observed. No dry ice growth was observed during flights 85-22 and 85-23. Also shown are the four wet/dry threshold curves calculated using the four different Van Fossen heat transfer coefficients (0.5% turbulence, rough and smooth surface; 3.5% turbulence, rough and smooth surface). These curves were calculated for the test cylinder diameter of 0.114 m and for the average flight airspeed of 71.4 m/s and the average exposure altitude of 1613 m.

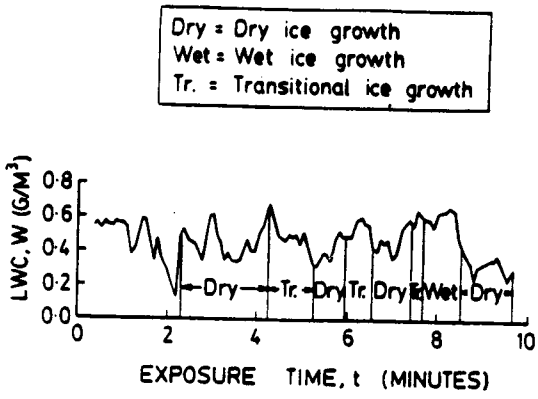


Fig. 19 Plot of liquid water content (measured by the Johnson-Williams probe) vs. exposure time for flight 85-24 showing typical fluctuations observed in natural icing conditions. Also shown are ultrasonic measured periods of wet, dry, and transitional ice growth.

Figs. 19 and 20 illustrate the considerable variations encountered in natural icing conditions, both during a particular flight and between flights conducted on different days. For example, the cloud temperature, liquid water content and droplet size were roughly comparable for flights 85-24 and 85-25. However, different ranges of wet and dry ice growth were observed, as indicated by the overlapping experimental wet and dry growth ranges at the same impinging liquid water content. The implication is that the heat transfer differed between the two flights, both at nominally similar icing conditions, but conducted on different days through different clouds.

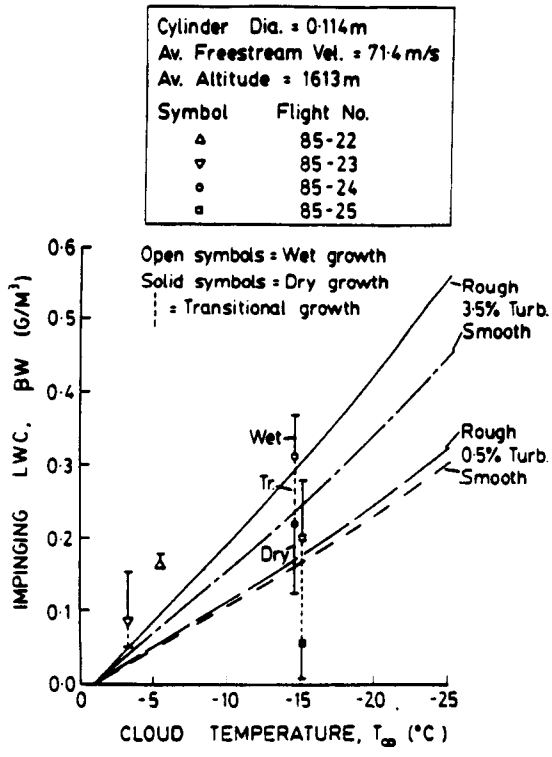


Fig. 20 Plot of impinging liquid water content vs. cloud temperature showing wet, dry and transitional ice growth regimes observed in flights and theoretical wet/dry threshold curves for four different heat transfer coefficients.

From the figure it can be seen that for flight 85-25 the high (3.5%) turbulence level heat transfer coefficients overpredict the observed heat transfer, based on the steady-state model analysis. For this flight the low (0.5%) turbulence level heat transfer coefficients appear appropriate since both of these coefficients correctly predicts the observed wet and dry regimes. The actual turbulence level applicable could be even less than 0.5% based on the location of the observed wet and dry growth regimes.

The experimentally observed ice growth regimes during flight 85-24 are consistent with the wet/dry threshold predicted by the 3.5% turbulence level, rough surface heat transfer coefficient. For this flight the low (0.5%) turbulence level models incorrectly predict wet growth for impinging liquid water content levels where dry growth was experimentally observed. Thus, in contrast to flight 85-25, the low turbulence level appears to be too low, and the 3.5% turbulence level model gives acceptable results.

The results of the natural icing tests also indicate the heat transfer occurring in natural icing conditions may vary from day to day despite similar icing conditions. One reason for this variation may be due to different icing cloud turbulence levels. Based on the limited amount of flight test data available it appears that in general the appropriate turbulence level for natural icing conditions is somewhat lower than the 3.5%+ level inferred for the icing research tunnel. This result is consistent with previous experimental comparisons between the icing research tunnel and flight. For these reasons, care should be taken in extrapolating the results of icing wing tunnel tests to "similar" natural icing cloud conditions.

4.0 MEASUREMENT OF ICE SHAPE BY AN ULTRASONIC ARRAY

Currently a comprehensive testing program is under way to measure and compare ice accretion in natural and artificial wind tunnel icing conditions. This ice-growth data base will enable a quantitative comparison to be made between flight and icing wind tunnel tests. In addition, this temporal and spatial ice accretion data will be used to support development and validation of analytic icing models and scaling laws. This section describes the experimental apparatus and testing procedure for these tests, and presents preliminary results of ice growth measurements made on an airfoil in natural icing conditions[4].

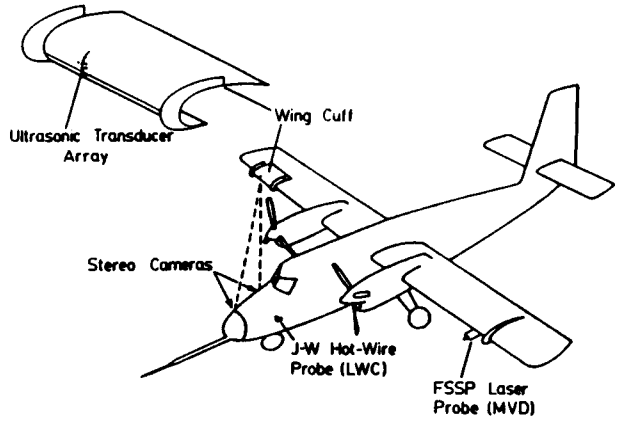


Fig. 21 Wing-cuff and ultrasonic array installation on NASA Lewis Twin Otter aircraft.

4.1 EXPERIMENTAL APPARATUS

For the natural icing flight tests a De Havilland DHC-6 airfoil section was instrumented with an array of 9 ultrasonic transducers, as shown in Fig. 21 The transducers were all 5 MHz, broadband, heavily damped transducers with 0.6 cm element diameters. The airfoil section was attached to the starboard wing of the NASA Lewis Icing Research Aircraft, a De Havilland DHC-6 Twin Otter. The airfoil section, or cuff, thus protruded approximately 3 in forward of the aircraft wing section.

The NASA aircraft was also equipped with a stereo camera system consisting of two 70 mm cameras mounted in the nose of the aircraft (see Fig. 21). The cameras produce stereo image pairs of the ice accretion on the wing cuff. These images are then photogrammetrically analyzed post-flight, and allow the ice accretion to be measured with a resolution of approximately ±0.03 in.

Fig. 22 schematically illustrates the ultrasonic system installation for the natural icing flight tests. The 9 ultrasonic transducers in the wing cuff were sequentially scanned using a multiplexing pulser/receiver unit. The pulse-echo signals were displayed on a broadband oscilloscope and a video camera focused on the oscilloscope screen provided a permanent record of these signals. Also within the camera's field of view were an electronic watch to provide time synchronization with other icing data recorded onboard the aircraft, and an LED display to indicate the current active transducer in the array.

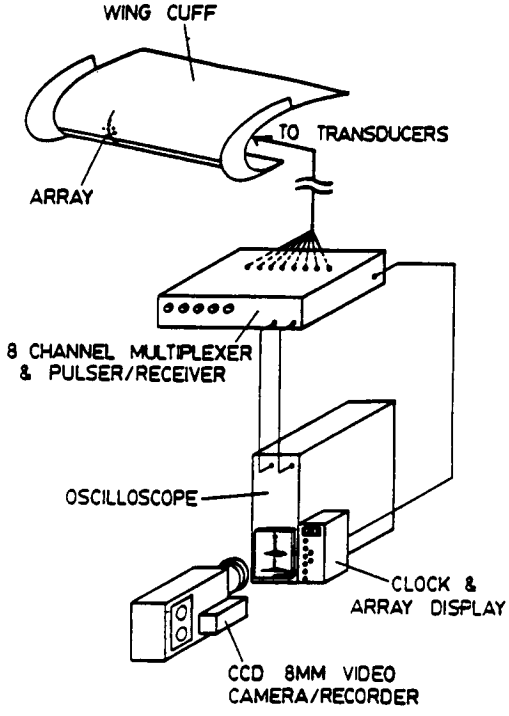


Fig. 22 Schematic of ultrasonic array equipment configuration.

4.2 TESTING

Since the instrumented wing cuff was permanently exposed on the aircraft wing, the ultrasonic system was activated from take-off to landing. Typically the multiplexing rate was set to approximately two seconds per transducer (i.e., each transducer was "active" for two seconds), so that four complete scans of the array were completed every minute.

In addition to the ultrasonic ice thickness measurements, stereo photographs of the wing cuff accretion were also taken during each encounter. However, adequate stereo photographs could not be obtained while the aircraft was inside the icing cloud, stereo photographs of the ice growth were therefore only taken outside the cloud at fairly widely spaced time intervals. Typically two or three separate stereo pairs were obtained documenting the ice growth during the icing encounter.

Since the wing cuff was not equipped with any ice protection system, in most cases it was possible to mechanically measure the final ice accretion on the wing cuff after landing using vernier calipers. Nine research flights were conducted with the ultrasonic system during the period February to March 1986. Due to the weather conditions during that period rime ice accretions were observed primarily.

4.2.1 Results

Fig. 23 shows a typical ice growth profile history obtained from the ultrasonic array. The data were from research flight 86-31 which had a cold average temperature of -10° C and a low liquid water content of 0.06 g/m^3. The data are presented in the form of successive ice profiles. These profiles were constructed by fairing a curve through the "point" thickness measurements from the array transducers. A total of six profiles are shown, with six minutes between each profile. The time at which the profiles were measured is indicated on the lower plot of the cloud liquid water content during the flight.

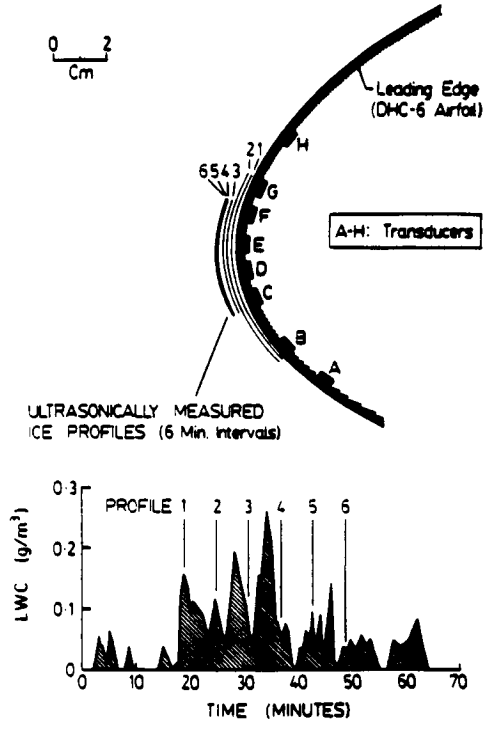


Fig. 23 Ultrasonically measured ice profiles for flight 86-31.

The ultrasonically measured profiles show the ice shape to be relatively conformable to the leading edge throughout the encounter. Thickness measurements from transducers B and G, located near the edges of the accretion, were not possible after the second profile. This was because the slope of the ice surface above these transducers, relative to the airfoil surface, became too large, reflecting the return echo away from the transducer and significantly reducing the received echo strength. Increasing the receiver gain in this situation would alleviate this problem; however, varying the receiver gain between transducers was not practical with the single multiplexed pulser/receiver used for these tests. Since a single "optimum" gain had to be used, this "edge" effect dropout of the echo signal was often unavoidable.

The ice profiles in Fig. 23 illustrate the non-uniform growth rate throughout the encounter. The first three profiles all show approximately equal growth, corresponding to the roughly constant average liquid water content during this period. The higher liquid water content in the interval between profiles 3 and 4 results in more growth, as evidenced by the larger profile spacing. Following profile 4 the liquid water content falls, and as a result profiles 5 and 6 show little further growth.

Fig. 24 illustrates the final ice shape accreted on the wing cuff at the completion of flight 86-31. Three separate measurements of the final ice profile are shown. The open circles represent thickness readings obtained from the stereo photographic analysis, while the crosses indicate measurements made with vernier calipers after landing. The final ultrasonic ice thickness measurements (from transducers C, D, E, and F) are shown as a solid line on the figure. The agreement between all three of these independent measurements is within 0.5 mm, with a final ice thickness of approximately 9 mm indicated.

5.0 CONCLUSION

In conclusion, ultrasonic pulse-echo measurement techniques are seen to offer the potential for the development of an operational ice detector, as well as providing a valuable tool for better understanding and documenting that ice accretion process.

ACKNOWLEDGMENTS

This work was supported by the National Aeronautics and Space Administration, the Federal Aviation Administration under the Joint University Program for Air Transportation, Grants NGL-22-009-640 and NAG-3-666, and the National Science Foundation Presidential Young Investigator Award Program. Wind tunnel and flight test facilities were provided by the NASA Lewis Research Center and the Boeing Commercial Airplane Company.

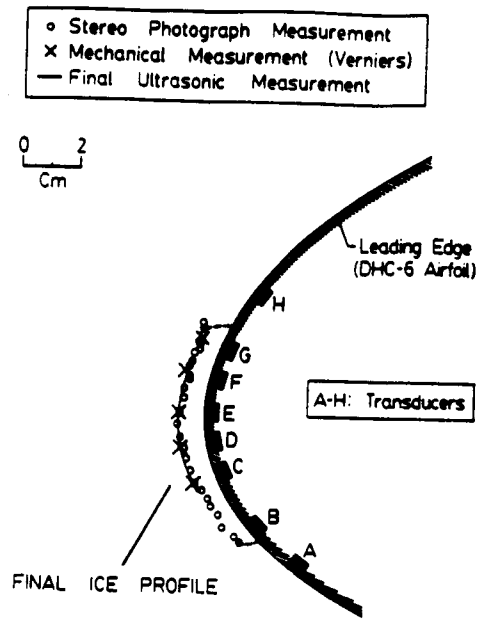


Fig. 24 Final ice profile for flight 86-31.

REFERENCES

1. Hansman, R.J. and Kirby, M.S., "Measurement of Ice Accretion Using Ultrasonic Pulse-Echo Techniques," *Journal of Aircraft*, Vol. 22, June 1985, pp. 530-535.

2. Van Fossen, G.J., et al., "Heat Transfer Distributions Around Nominal Ice Accretion Shapes Formed on a Cylinder in the NASA Lewis Icing Research Tunnel," AIAA-84-0017, January 1984.

3. Hansman, R.J. and Kirby, M.S."Comparison of Wet and Dry Growth in Artificial and Flight Icing Conditions" *Journal of Thermophysics and Heat Transfer*, Vol 1, July 1987, pp.215-227.

4. Hansman, R.J., Kirby, M.S., McKnight, R., and Humes R., "In-Flight Measurement of Airfoil Icing Using an Array of Ultrasonic Transducers," AIAA-87-0178, 1987.

Investigation of Surface Water Behavior During Glaze Ice Accretion*

R. John Hansman Jr.* and Stephen R. Turnock†
Massachusetts Institute of Technology, Cambridge, Massachusetts

A series of experimental investigations that focused on isolating the primary factors that control the behavior of unfrozen surface water during glaze ice accretion were conducted. Detailed microvideo observations were made of glaze ice accretions on 2.54 cm diam cylinders in a closed-loop refrigerated wind tunnel. Distinct zones of surface water behavior were observed; a smooth wet zone in the stagnation region with a uniform water film, a rough zone where surface tension effects caused coalescence of surface water into stationary beads, and a zone where surface water ran back as rivulets. The location of the transition from the smooth to the rough zone was found to migrate towards the stagnation point with time. Comparative tests were conducted to study the effect of the substrate thermal and roughness properties on ice accretion. The importance of surface water behavior was evaluated by the addition of a surface tension reducing agent to the icing tunnel water supply, which significantly altered the accreted glaze ice shape. Measurements were made to determine the contact angle behavior of water droplets on ice. A simple multizone modification to current glaze ice accretion models was proposed to include the observed surface roughness behavior.

Nomenclature

A	= surface area, m^2
C_i	= specific heat capacity of ice, J/kg K
C_p	= specific heat capacity of air, J/kg K
C_w	= specific heat capacity of water, J/kg K
D	= diffusion coefficient of water vapor in air, m^2/s
F	= force, kg m/s^2
h	= local convective heat-transfer coefficient, W/m^2K
k	= thermal conductivity of air, W/m K
L_f	= latent heat of fusion of water, J/kg
L_s	= latent heat of sublimation of water, J/kg
L_v	= latent heat of vaporization of water, J/kg
$\dot{M}''$	= local mass flux/time, kg/m^2s
$\dot{Q}''$	= local heat flux/time, W/m^2
r	= recovery factor, 0.875(-)
T_{surf}	= equilibrium surface temperature, °C
T_∞	= cloud temperature, °C
t	= icing time, s
V_∞	= freestream velocity, m/s
W	= cloud liquid water content, g/m^3
β	= local collection efficiency (-)
$\Delta\theta$	= contact angle hysteresis, deg
ΔT_∞	= cloud supercooling = T_∞(°C)
θ	= contact angle, deg
θ_a	= advancing contact angle, deg
θ_r	= receding contact angle, deg
$\rho_{v,surf}$	= saturated vapor density over surface, kg/m^3
$\rho_{v,\infty}$	= saturated vapor density in cloud, kg/m^3
σ	= surface tension, kg/m s^2

Received Nov. 26, 1987; presented as Paper 88-0015 at the AIAA 26th Aerospace Sciences Meeting, Reno, NV, Jan. 11-14, 1988; revision received Feb. 8, 1988.

*Associate Professor, Aeronautics and Astronautics. Member AIAA.

†Research Assistant, Aeronautics and Astronautics; currently at The University of Southampton, England.

*J. Aircraft, vol. 26, no. 2, 1989.

I. Introduction

THE behavior of unfrozen water on an accreting glaze ice surface can directly and indirectly influence the shape of the resulting ice accretion. The surface transport behavior of the unfrozen water prior to freezing has a direct impact on the ice shape due to its effect of redistributing the impinging water mass. The surface water behavior also indirectly influences the ice accretion through its impact on surface roughness, which modifies the local convective heat transfer. The local heat transfer is the controlling factor in wet surface glaze ice accretion, where there is sufficient impinging water such that ice accretion rate is limited by the ability to remove latent heat of fusion from the surface.

Current ice accretion models ignore the specific details of unfrozen surface water behavior during glaze accretion. The transport behavior and surface roughness are both treated in a simplistic or heuristic manner. This simplistic evaluation of surface roughness and transport behavior, which omits consideration of the surface physics, is considered to be a contributing factor in the poor agreement between current glaze ice accretion models and expeimental results.[1]

This paper describes a series of experimental investigations focused on isolating the primary factors that control surface water behavior. The experimental investigations include measurement of the contact angle and resistance to motion of water on an ice surface, photographic observations of surface roughness during glaze ice accretion on cylinders, differential comparisons of substrate thermal and surface properties on ice accretion, and the influence of surface tension on glaze ice accretion. The results indicate the importance of surface water behavior to the icing process in the glaze ice regime. A simple modification to existing glaze ice models is also proposed.

II. Ice Accretion Mechanisms Influenced by Surface Water Behavior

The primary ice accretion mechanisms that are influenced by the behavior of unfrozen surface water are the convective heat transfer, which is strongly related to the surface roughness,[2] and the mass transfer at the surface, which is partly determined by the water runback characteristics of the surface. Both the

heat and the mass transfer will be discussed briefly in the context of ice accretion models to identify the importance of surface water behavior on the ice accretion process.

Mass Transfer

The principal modes of liquid water mass transfer on an accreting ice surface are shown schematically in Fig. 1, which depicts mass flux into and out of a control volume at the ice surface. The primary source of liquid water is the impingement of supercooled droplets. The impinging liquid water mass flux $\dot{M}''_{imp}$ is linearly related to the ambient liquid water content W, the local droplet collection efficiency β, and the freestream velocity V_∞:

$$\dot{M}''_{imp} = \beta W V_\infty \tag{1}$$

The double prime superscript is used to indicate that the quantity is defined per unit area of the icing surface.

Liquid water may also enter the control volume through mass flow along the surface ($\dot{M}_{in}$ in Fig. 1). Liquid water leaves the control volume through freezing at the ice water interface $\dot{M}''_{freeze}$, and surface flow out of the control volume $\dot{M}_{out}$. Some mass also leaves the control volume due to evaporation, but this is generally small in the icing problem and will be neglected in the following analysis. The freezing mass flux $\dot{M}''_{freeze}$ is determined by the ability to convect the latent heat resulting from the water ice phase transition away from the surface. This will be discussed more fully in the following section.

In steady state, the mass flux into and out of a control volume with surface area A must balance

$$A \cdot \dot{M}''_{imp} + \dot{M}_{in} = A \cdot \dot{M}''_{freeze} + \dot{M}_{out} \tag{2}$$

For rime icing conditions at cold temperatures, the convective heat transfer is sufficient to freeze all of the impinging mass flux. In this case, there is no surface flow, and Eq. (2) reduces to

$$\dot{M}''_{imp} = \dot{M}''_{freeze} \tag{3}$$

In glaze icing conditions, there is insufficient heat transfer to freeze all of the impinging mass flux. In this case, the surface flow terms in Eq. (2) have to be considered. In current ice accretion models, all unfrozen water is assumed to run back out of the control volume, while unfrozen water from the next upstream element is assumed to flow into the control volume.[1]

The assumption of uniform runback is quite reasonable for icing conditions where a thin water film covers the glaze ice surface. However, recent NASA high-speed photographic studies indicate that other runback mechanisms may exist.[3,4] In some of the NASA studies, surface water was observed to coalesce into beads that remained stationary on the accreting ice surface. An example is shown in Fig. 2. The liquid beads were often surrounded by regions of otherwise dry ice surface, and no runback was observed. The NASA observations suggest two potential scenarios for surface water flow in the presence of water beads. In one scenario, the water beads will grow until the increased heat transfer from the rough surface becomes sufficient for the freezing mass flux to equal the impinging mass flux and any surface flow into the control volume. In this case, no water would leave the control volume, and Eq. (3) would reduce to

$$A \cdot \dot{M}''_{freeze} = A \cdot \dot{M}''_{imp} + \dot{M}_{in} \tag{4}$$

In the alternate scenario, the water beads would grow through coalescence and impingement until external dynamic pressure and shear forces caused the droplets to slide along or be torn from the surface. In this case, it is unlikely that the drop would move into the next control volume, and the assumption of surface flow continuity implicit in current mass balance models does not hold. This may explain experimental observations of reduced total ice accretion at temperatures near 0°C.[5]

It should be noted that both scenarios described above may occur depending on the particular temperature and liquid water conditions. It should also be noted that uniform water films were also observed in the NASA experiments, particularly near the stagnation region.

Heat Transfer

The thermodynamic analysis presented in this paper for an accreting ice surface follows the earlier work by Olsen et al.[5] and others,[6,7] and is commonly employed in current ice accretion models.[1,8] The principal modes of energy transfer associated with the icing surface are depicted schematically in Fig. 3. Heat is added to the surface, primarily from the latent heat of

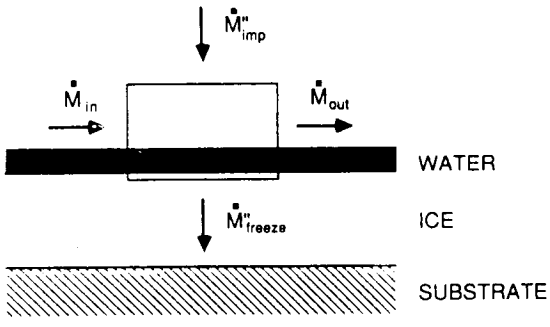


Fig. 1 Control volume mass balance for a wet surface.

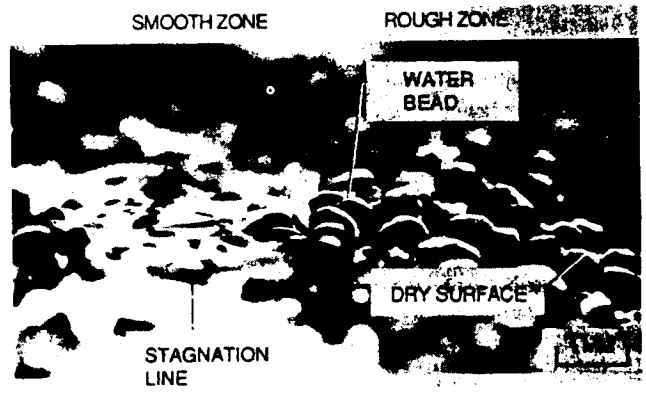


Fig. 2 Grazing angle photograph of a glaze ice surface (courtesy of William Olsen, NASA Lewis Research Center).[4]

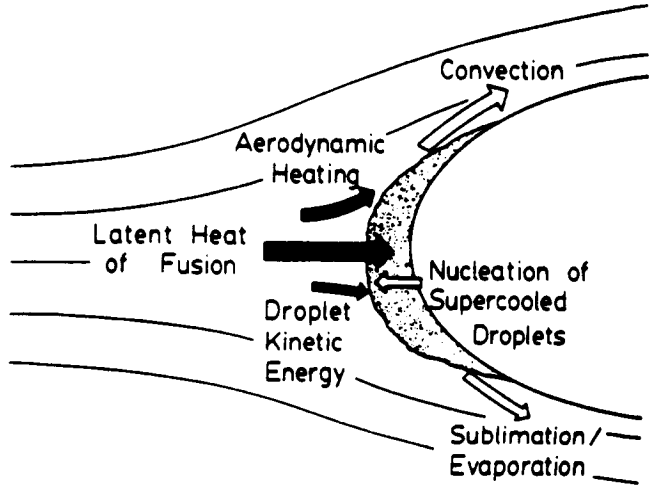


Fig. 3 Modes of energy transfer for an accreting ice surface.

fusion released as the droplets freeze, but also from aerodynamic heating and, to an even smaller extent, from the kinetic energy of the droplets impacting the surface. Heat is removed from the surface primarily by convection, and to a lesser degree by sublimation (when the surface is dry) or evaporation (when the surface is wet). In addition, heat is absorbed from the surface as the supercooled droplets impinge and warm to 0°C.

The assumption of steady state requires that the rate at which energy is added to the control volume equals the rate at which it is removed, i.e.,

$$\dot{Q}''_{in} = \dot{Q}''_{out} \tag{5}$$

where $\dot{Q}''_{in}$ and $\dot{Q}''_{out}$ represent the energy added to and removed from the control volume per unit area per unit time. Equation (2) may be expanded into its component energy terms as

$$\dot{Q}''_{in} = \dot{Q}''_{freezing} + \dot{Q}''_{aero\ heating} + \dot{Q}''_{droplet\ kinetic\ energy} \tag{6}$$

$$\dot{Q}''_{out} = \dot{Q}''_{conv} + \dot{Q}''_{sub/evap} + \dot{Q}''_{drop\ warming} \tag{7}$$

At steady state it is assumed that the ice surface achieves a locally uniform equilibrium temperature T_{surf}. Conduction into the ice is assumed to be zero, and chordwise conduction between adjacent control volumes is neglected. With these assumptions, the component heat terms of Eqs. (6) and (7) may be written as

$$\left.\begin{aligned} \dot{Q}''_{freezing} &= \dot{M}''_{freeze}[L_f + C_i(0°C - T_{surf})] \\ \dot{Q}''_{aero\ heating} &= \frac{rh\,V_\infty^2}{2C_p} \\ \dot{Q}''_{droplet\ kinetic\ energy} &= \frac{\dot{M}''_{imp}\,V_\infty^2}{2} \end{aligned}\right\} \dot{Q}''_{in} \tag{8a}$$

$$\left.\begin{aligned} \dot{Q}''_{conv} &= h(T_{surf} - T_\infty) \\ \dot{Q}''_{sub/evap} &= \frac{hDL}{k}(\rho_{v,surf} - \rho_{v,\infty}) \\ \dot{Q}''_{droplet\ warming} &= \dot{M}''_{imp}\,C_w\Delta T_\infty \end{aligned}\right\} \dot{Q}''_{out} \tag{8b}$$

Note that both of the primary heat dissipation terms $\dot{Q}''_{conv}$ and $\dot{Q}''_{evap}$ contain the local convective heat-transfer coefficient h. Also note that the $\dot{Q}''_{freezing}$ term that dominates the heat flux into the control volume contains the freezing mass flux $\dot{M}''_{freeze}$. This implies that the local ice accretion rate is strongly coupled to the local convective heat-transfer coefficient. For wet surfaces, the ice accretion rate is essentially limited only by the capability to transfer the latent heat of the freezing water away from the surface.

The convective heat-transfer coefficient h has been empirically observed to be strongly dependent on the ambient surface roughness.[2] Current ice accretion models either empirically generate local convective heat-transfer distributions to match natural ice shapes,[9] or assume a uniform roughness and use integral boundary-layer techniques to derive heat-transfer distributions.[10] In these cases, the surface roughness is characterized by an equivalent sand-grain roughness, which is adjusted to match experimental ice accretions.

The surface water behavior is a primary factor in the evolution of roughness on an accreting glaze ice surface. The surface roughness determines the heat-transfer behavior, which ultimately determines the ice accretion rate under glaze conditions. Understanding the generation of surface roughness by the microphysical surface water behavior is, therefore, a key component to understanding the glaze ice accretion process.

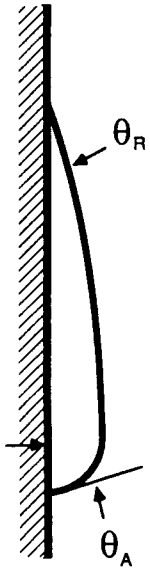


Fig. 4 Example of contact angle hysteresis in a droplet on a vertical wall.

III. Microphysics of Surface Water Behavior

Forces Influencing Surface Water Behavior

The microphysical behavior of water on an aircraft surface is controlled primarily by the relative strength of the surface tension, and by aerodynamic and body forces. Surface tension forces tend to minimize the surface area of the fluid, causing the water to coalesce into beads or rivulets. They also act to oppose motion of fluid along the surface. Body forces such as gravity or centripetal acceleration act on the entire fluid bulk. Aerodynamic forces result from pressure gradients and shear stress at the water-air interface.

Contact Angle

An important parameter in surface fluid behavior is the contact angle θ. It is defined as the angle the fluid-vapor interface makes with an underlying substrate. The nominal contact angle is a property of the specific gas-liquid-solid combination, and may vary with temperature. The wettability, or tendency of a fluid to spread on a particular surface, is inversely related to the contact angle θ. In general, for contact angle values of less than 10 deg, the surface is considered wettable.

If the fluid is subject to an external force such as gravity, the contact angle may vary, depending on the direction of the force relative to the contact line. This hysteresis effect can be visualized in the schematic view of a drop sliding down a vertical surface, shown in Fig. 4. On the lower surface of the drop the advancing line of contact causes an increase in the contact angle to its maximum allowable value θ_a. Conversely, the receding edge of the drop remains at the lowest allowable value of the contact angle θ_r. The contact angle hysteresis $\Delta\theta$ is the difference between the advancing and receding contact angles:

$$\Delta\theta = \theta_a - \theta_r \tag{9}$$

The contact angle hysteresis $\Delta\theta$ is a property of the specific gas-liquid-solid combination, and will tend to increase with surface roughness. The resistance to motion of the liquid-solid line of contact can be related to the value of the nominal contact angle and its hysteresis.[11] In general, increasing values of θ and $\Delta\theta$ imply increased resistance to motion.

Modes of Surface Water Behavior

Water on an aircraft surface tends to behave in one of three distinct modes depending on the water flow rate, aerodynamic forces, and the wettability of the surface. For high flow rates and wettable surfaces, the water will tend to coat the entire surface with a uniform film. For lower flow rates, or higher contact angles, there is insufficient water to coat the entire surface. Once drying begins, surface tension forces will tend to

coalesce the available water into either rivulets or water beads.

Water beads are generally associated with low flow rates or unwettable surfaces such as wax. Equation (10) represents the overall force balance for a small stationary water bead. As long as the contact line force is greater than the sum of the aerodynamic and body forces, the bead will remain stationary.

$$F_{\text{contact}} \geq F_{\text{aero}} + F_{\text{body}} \tag{10}$$

The contact line force F_{contact} is the total force resulting from the contact line resistance to motion described in the preceding section. Both the contact line force and the aerodynamic force are roughly proportional to the beads surface area, whereas the body force is proportional to the bead volume. This implies that, as the bead size is increased, there will be maximum stationary bead volume for a given set of surface, flow, and gravitational conditions above which the bead will move.

IV. Measurement of Contact Angle Between Ice and Water

As discussed in Sec. III, the contact angle θ and hysteresis $\Delta\theta$ are important parameters in the behavior of water on an ice surface. There is, however, very little data available on θ and $\Delta\theta$ for water on ice. This is partly due to experimental difficulties. Most investigations of the surface properties of supercooled water have concentrated on measurements of surface tension σ.[12] In order to obtain preliminary working values of θ and $\Delta\theta$ for water on ice, a series of simple experiments were conducted.

Experimental Setup

The apparatus employed in these investigations is shown schematically in Fig. 5. A smooth layer of ice, approximately 8 mm thick, was formed from distilled water on a metal plate that could be oriented at a variety of angles with respect to the horizontal. Precooled droplets of distilled water were placed on the ice surface by a syringe, and their shapes were recorded by a CCD microvideo camera oriented to view the ice surface at a grazing angle. By use of a dark background and a diffuse light source, a sharp high-contrast image of the droplet at the ice-water interface could be obtained. This enabled contact angles to be measured to an accuracy of approximately ± 5 deg. The ice surface temperature was monitored by an iron-constantan thermocouple mounted in direct contact with the ice surface. The output of the thermocouple was presented on a digital display within the field of view of the video camera. This produced a simultaneous record of contact angle and surface temperature valid to approximately ± 1°C.

Measurement of the contact angle at various temperatures was obtained by initially cooling the iced metal plate in a cold box, to a temperature below the target measurement temperature. The plate was then removed from the cold box and mounted horizontally within the field of view of the camera. As the plate warmed to the target temperature, droplets were placed on the surface and their shapes were recorded. Although the droplets would begin to freeze shortly after placement, the initial contact angle was preserved in the frozen drop, and any transient effects could be observed in the video record. In this manner, values of the contact angle of water on ice could be obtained for subfreezing temperatures.

The contact angle hysteresis measurements were made utilizing a similar procedure. However, the iced plate was set to an angle of 30 deg with respect to the horizontal. Each droplet's volume was increased by syringe injection until motion began; at this instant the advancing and receding angles and the surface temperature were measured from the video recording.

Results

The observed dependence of contact angle θ and contact angle hysteresis $\Delta\theta$ with temperature are shown in Figs. 6 and 7, respectively. Because of the low magnitude, it was not possible to obtain accurate contact angle measurements at temperatures above -4°C.

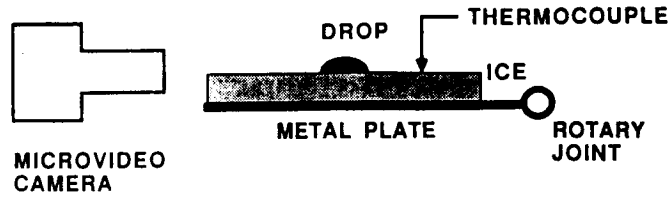


Fig. 5 Schematic of contact angle measurement setup.

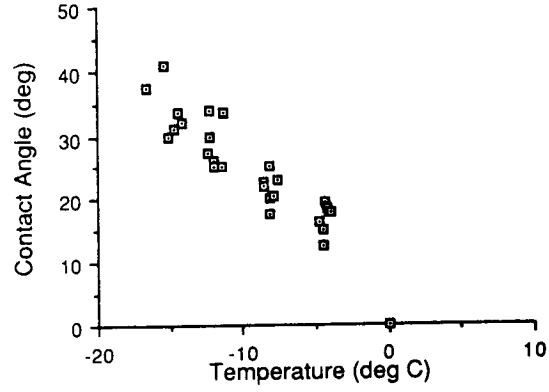


Fig. 6 Contact angle temperature variation for water on ice.

However, at the freezing point (0°C), ice and water are in equilibrium so that the ice surface must be perfectly wettable, implying that both θ and $\Delta\theta$ must be zero. These points are included in the contact angle and contact angle hysteresis plots. They are consistent with the experimentally observed values and allow interpolation between -4°C and 0°C.

Both the contact angle and hysteresis exhibit a strong nonlinear variation with temperature, particularly in the vicinity of the freezing point (0°C). These results imply that the wettability of the ice surface decreases greatly as the ice surface cools below freezing. For warm ice surfaces (near 0°C) water will tend to spread, and the ice-water contact line will have a low resistance to motion. For colder ice temperatures, water will tend to bead into droplets, which require a higher force to initiate fluid motion.

The strong contact angle temperature dependance indicates the potential importance of thermal gradients to the development of ice surface roughness. Small variations in surface temperature can restrict the mobility of water. This effect is thought to be the cause of the stable nature of the surface water beads observed by Olsen and Walker[3,4] and shown in Fig. 2. The dry ice surface around the beads can be cooled to temperatures well below 0°C by convective heat transfer. This cold dry surface, therefore, imposes a barrier to water flow away from the bead. Any impinging water that strikes the dry surface will quickly be frozen due to the low temperature. Any cloud droplets that strike the water will be trapped within the bead, while glaciation occurs at the lower surface of the bead. In this manner, the ice thickness can increase, while the bead is observed to remain stationary.

V. Experimental Observations of Ice Surface Roughness on Cylinders

Experiment Setup

Detailed photographic observations of the behavior of surface water and formation of ice roughness on cylinders during glaze ice accretion were made in the Data Products of New England Icing Wind Tunnel. The experimental setup is shown

in Fig. 8. The tunnel was a closed-loop refrigerated system with a 6-in.-square plexiglass test section. The test article was a cylinder that horizontally spanned the test section. Scale reference for the photographic studies was provided by a grid mounted on a thin splitter plate at the midplane of the test section. An 8 mm CCD microvideo (camera A) with a macrolens for magnification was used to obtain a grazing angle view of the ice accretion. The camera was focused at the stagnation region near the center of the test section. An additional video camera (camera B) was mounted above and slightly upstream of the cylinder to provide a view of the ice accretion looking normal to the cylinder surface. Secondary lighting was provided to obtain a clear video record of the accreting ice surface.

The tunnel and icing cloud parameters were set to yield a variety of glaze ice conditions. The air velocity was 150 knots for all tests. A range of freestream air temperatures between -9 and -4°C was used. The centerline liquid water content W was varied from 0.7 to 1.2 g/m^3. The icing cloud for the conditions had a nominal mean volumetric diameter (MVD) of 30 μ. The standard exposure time was 3 min. The tunnel had previously been calibrated using an indirect method. There is, therefore, some uncertainty in the liquid water content and MVD values.

The liquid water content was observed to vary somewhat across the test section with a reasonably constant valve in the central 2 in. as measured by rime ice accretion calibration runs. Photographic observations were, therefore, focused on the centerline region where the ice deposit was uniform and the liquid water content calibration was valid.

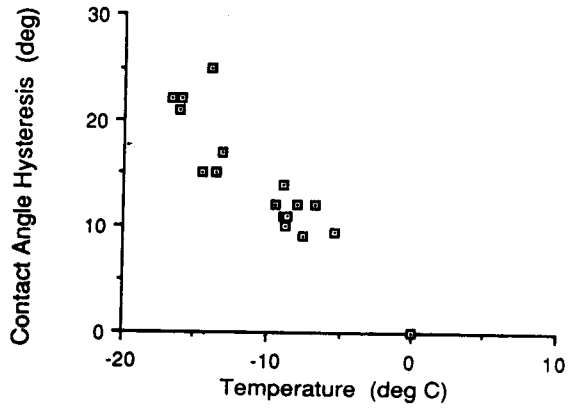


Fig. 7 Contact angle hysteresis temperature variation for water on ice.

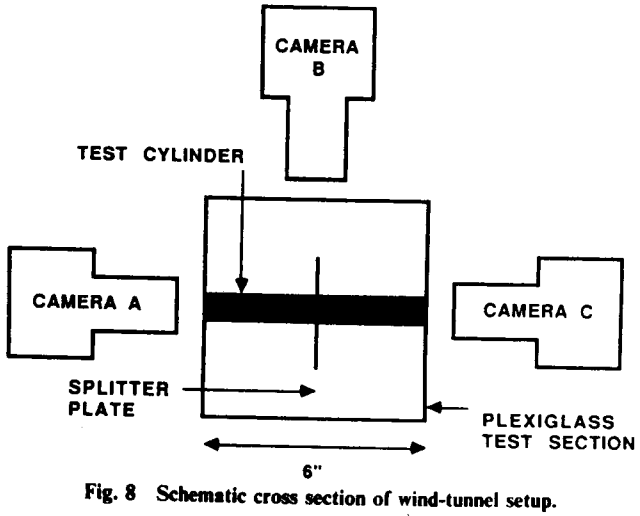


Fig. 8 Schematic cross section of wind-tunnel setup.

Results

Figure 9 shows a representative final glaze ice shape for an exposure temperature of -4.5°C and a liquid water contact of 1.0 g/m^3. During the accretion, the following three distinct types of ice surface behavior were observed, each having a characteristic roughness and identifiable boundaries.

1) *Smooth Zone.* Close to the stagnation point during exposure to the icing cloud, the surface was observed by light reflection techniques to be uniformly wet with a thin film of water at warm temperatures. The surface in this regime was smooth, with no distinct roughness. The ice was translucent within this zone.

2) *Rough Zone.* At some point downstream, there was a sudden transition to a significantly rougher surface. Within this zone, there appeared to be insufficient water to maintain a uniform film. Surface tension forces dominated the water surface behavior. Runback did not occur; rather, the water tended to coalesce into the water beads first observed by Olsen and Walker.[3,4] The scale length of the roughness was typically on the order of 1 mm. The transition between the rough and the smooth zones can be clearly seen in the grazing angle photograph shown in Fig. 2.

Inasmuch as there was a distinct boundary between the smooth and rough zones, this position could be easily identified on the grazing angle video recording. The angular position of this boundary is plotted in Fig. 10 for cold (-9°C) and warm (-4.5°C) conditions at a liquid water content of 1.0 g/m^3. The boundary started at approximately 50 deg and propagated rapidly towards the stagnation region. The repeatable nature of the smooth-rough transition's propagation toward the stagnation region implies a clear underlying physical mechanism

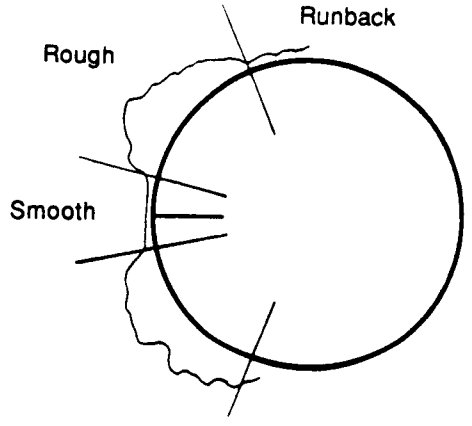


Fig. 9 Typical final glaze ice shape showing distinct roughness zones, $T = -4.5$°C, $W = 1.0$ g/m^3.

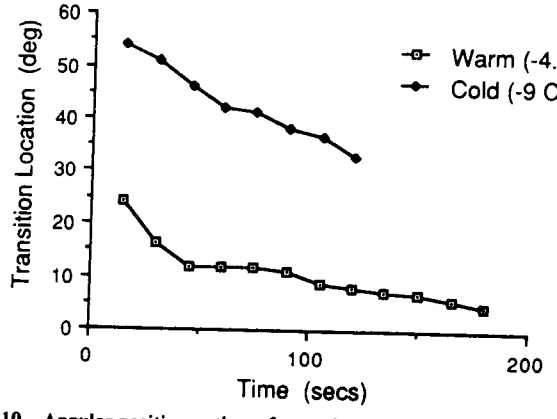


Fig. 10 Angular position vs time of smooth-rough transition location for two air temperatures.

for the transition between the surface water behavior in the smooth and rough zones.

The ice accretion rate was observed to be enhanced in the rough zone as compared with the smooth zone. This may be seen in the ice profiles of Fig. 11 taken from the video at 30, 90, and 150 s after exposure to the icing cloud. The enhanced accretion rate is thought to be due to increased heat transfer resulting from the greater surface roughness in this zone. In all of the cases observed in this study, the ice horns characteristic of glaze ice accretion were found within the rough zone.

3) *Runback Zone.* At warm temperatures a third zone was observed aft of the rough zone. This region was characterized by areas of ice interspersed with uniced surface. This ice was observed to form during an initial transition period after cloud exposure. The ice was translucent, and quite often frozen rivulets could be discerned. In warm conditions and high liquid water contents, the surface water was observed to initially runback and then stagnate at the point of flow separation. This water then slowly froze as rivulets or as large coalesced water cells. Once ice began to form in the upstream rough zone, no additional surface water was supplied to the runback zone, and the ice surface remained constant.

VI. Observations of the Effect of Substrate Properties on Surface Water Behavior

Experimental Setup

Two properties of the initial uniced surface that can affect the surface water behavior are 1) thermal characteristics (i.e., conductivity, specific heat capacity), and 2) roughness of the uniced surface. By exposing two test articles (identical except for either their thermal characteristics or initial surface roughness) to the same icing cloud, it was possible to assess the individual importance of each parameter. Comparative tests of this type were carried out in the icing wind-tunnel setup described in Sec. V. For these experiments, the test article was composed of two 3-in.-long, 1-in.-diam cylinders of different material composition, separated by a splitter plate. As shown in Fig. 8, each half of the test article was viewed by a grazing angle microvideo camera (cameras A and C). A third camera (B) was mounted above and upstream of the article to simultaneously record the differences between the two cylinders.

Two cylinders were used in the thermal comparison experiments: a copper tube with a 1/16-in. wall thickness to investigate fast thermal response, and a solid plexiglass rod to investigate slow thermal response. To remove the possible influence of surface chemistry effects, each cylinder was covered with a single coat of acrylic paint.

For the roughness experiment, two cylinders were used. Both were manufactured from solid aluminum rods. One was extremely smooth with a polished finish obtained using '0000' emery paper. The rough cylinder had a repeatable surface pattern that was produced by knurling the cylinder on a lathe. The knurling process produces a pattern of trapezoidal surface elements. For the cylinder used in these experiments, the elements had a height of 0.8 mm, a width of 1 mm, and a length of 2.7 mm.

Results

Thermal Comparison

Figure 12 shows a typical comparison of the final ice shapes obtained after 3 min exposure for the copper tube and the plexiglass rod. This comparison was run at a temperature of -5.5°C and a liquid water content of 0.95 g/m^3. Although there is not a great deal of difference between the final accretions, the glaze ice horns on the copper cylinder were more sharply defined than on the plexiglass, which had a relatively flat front surface and was slightly thicker in the stagnation region. The angular position of the rough-smooth transition boundary is shown as a function of exposure time for these accretions in Fig. 13. For the plexiglass cylinder, the smooth zone shrinks much more rapidly than for the copper. The ice surface on the plexiglass cylinder became uniformly rough after 45 s of exposure. The enhanced roughness in the stagnation region, after this time, accounts for the thicker ice in the stagnation region and the relatively flat front ice surface observed in Fig. 12.

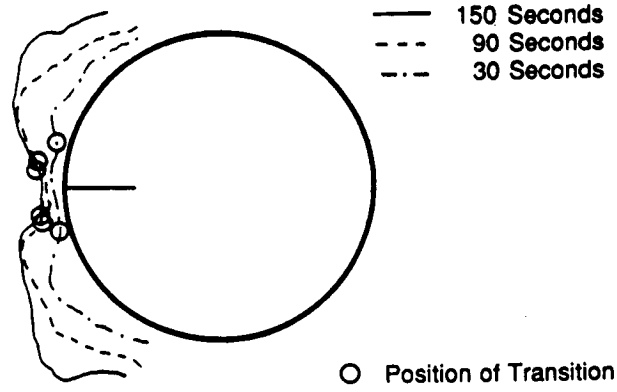


Fig. 11 Time development of glaze ice shapes, $T = -4.5$°C, $W = 1.0$ g/m^3.

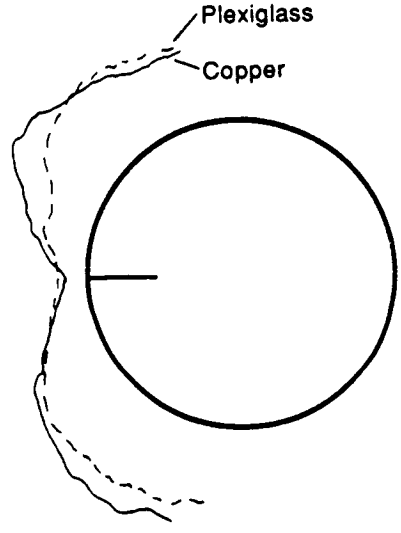


Fig. 12 Typical final ice shape comparison for copper and plexiglass cylinders, $T = -5.5$°C, $W = 0.95$ g/m^3.

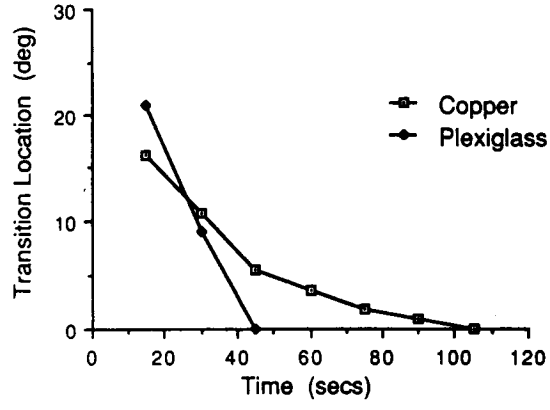


Fig. 13 Angular position of smooth-rough transition vs time for copper and plexiglass cylinders.

The water surface behavior after initial exposure to the icing cloud was observed to differ between the copper and plexiglass test articles. For the copper cylinder, there was relatively little surface water flow prior to freezing, and the initial ice surface was relatively smooth. For the plexiglass, droplet coalescence and runback were observed to occur prior to the initial ice formation. This resulted in a significantly rougher initial ice surface, which is thought to be the cause of the difference in accretion behavior observed between the plexiglass and copper test articles.

In general, the differences in ice accretion behavior due to substrate thermal properties were observed to be most significant during the initial phase of ice accretion. As the accretion grows, the effect of initial conditions begins to wash out. After extended icing encounters, the accretion will tend toward a shape that is controlled by the environmental parameters and is independent of substrate properties.

Uniced Surface Roughness

A typical comparison of final ice profiles for the polished (smooth) and knurled (rough) cylinders is shown in Fig. 14. The ice horns on the knurled cylinder were broader and further off the stagnation line than the horns on the polished cylinder. This is thought to be due to enhanced heat transfer in the horn regions during the initial phase of the ice accretion. A slight but repeatable asymmetry was observed in the polished ice growth. This is thought to be due to the smooth surface on the polished cylinder, which allowed gravitational forces to influence the initial surface mass flux, resulting in a larger ice horn on the lower surface.

The zones of smooth and rough ice growth discussed in Sec. V were observed in both the polished and knurled cases. Figure 15 plots the angular position of the rough-smooth transition vs time. There is some difference between the two cases, with the smooth zone being smaller for the knurled cylinder after approximately 80 s exposure time. An important observation was that the trapezoidal element pattern of the knurled cylinder was clearly apparent in the ice accretion within the rough zone. This implies that the roughness geometry propagates in the ice accretion within the rough zone. In the absence of uniced surface roughness, water beading mechanisms appear to be the source elements for the roughness geometry.

VII. Surface Tension Effects on Ice Accretion

Experimental Setup

The importance of surface water behavior was further investigated by varying the surface tension of the water and studying the effect on the resulting ice accretion. The water surface tension was varied by the addition of a surfactant (Photoflo 200), which reduced the surface tension by approximately a factor of 2, while leaving its bulk properties (density, freezing temperature) unchanged.

The icing wind-tunnel setup was similar to that discussed in Sec. V. Photoflo was added at a dilution of 1:200 to the icing cloud water supply. Icing tunnel conditions were matched as closely as possible between runs carried out with and without the surfactant added. Because the spray system was not specifically calibrated with Photoflo, some uncertainty exists on liquid water content and MVD values. Based on splitter plate ice accretion, this uncertainty is estimated to be on the order of 5%. Photoflo runs were conducted at a liquid water content of 1 g/m^3 and for temperatures between -9 and -4°C.

Results

Final glaze ice profile comparisons between a normal icing cloud and the Photoflo modified cloud exposures are shown in Fig. 16 for the copper cylinder at -4.5°C. The addition of Photoflo resulted in an opaque white ice with a smaller scale surface roughness and a significantly different profile. The horns were more pronounced and closer to the stagnation point. At the lower temperatures, the glaze ice horn has be-

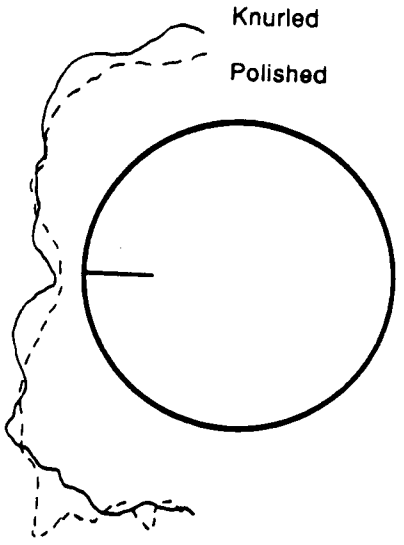


Fig. 14 Typical final ice shape comparison for polished and knurled cylinders, $T = -4.5$°C, $W = 1.0$ g/m^3.

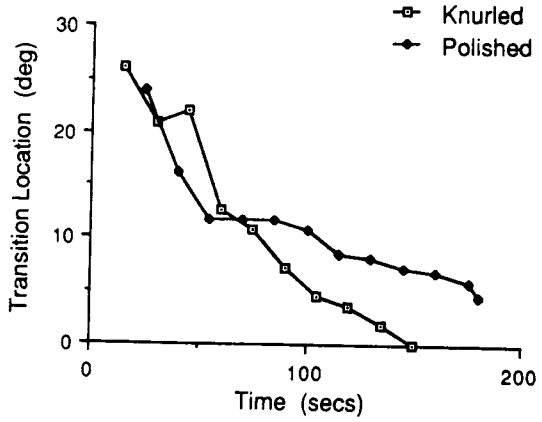


Fig. 15 Angular position of smooth-rough transition vs time for polished and knurled cylinders.

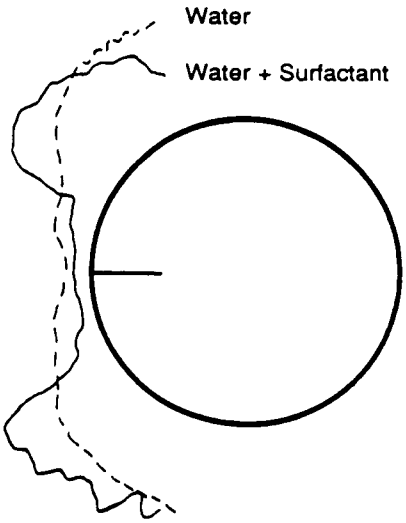


Fig. 16 Typical final ice shape comparison for cloud water supply with and without surfactant, $T = -4.5$°C, $W = 0.95$ g/m^3.

come so pronounced that it no longer collected all of the impinging water. A second horn was, therefore, able to develop behind it.

The different and unusual ice accretions obtained at reduced values of surface tension indicate that surface water behavior is, indeed, an important factor in the glaze ice accretion process and must be considered in physically realistic models.

VIII. Implications for Ice Accretion Modeling

Based on the observations of surface water behavior during glaze ice accretion, a relatively simple modification to the existing ice accretion model is proposed, which may improve the current model's accuracy within the glaze ice regime. In this proposed "multi-zone" model, the accreting ice surface is divided into two or more discrete zones that have varying surface water behavior and surface roughness. This is in contrast to current techniques, which assume that the surface has uniform roughness and surface water runback.

In the simplest version of the model, the surface is divided into two zones. There is a smooth wet zone centered about the stagnation region where thin film runback occurs, and the existing Messinger-type models[6] appear to be valid. In this region the heat transfer would be that for a smooth surface. The remaining ice surface would consist of a "rough" zone where surface tension effects dominate the surface water behavior, and the characteristic water beads appear. The heat transfer within this region would be adjusted by incorporation of roughness correction factors appropriate to the bead dimensions.

Two primary areas need to be addressed prior to implementation of a two-zone model. They are: 1) determining the smooth-rough transition locations, and 2) determining the appropriate value of equivalent sand-grain roughness to be used in calculating the rough zone heat-transfer coefficient.

The work described in Sec. V indicates that there are distinct physical mechanisms that control the position of the smooth-rough transition. However, further work is necessary to parametrically define the location of the smooth-rough transition. In addition, the factors that control the generation of roughness elements within the rough zone need to be investigated in the context of their impact on the local convective heat-transfer coefficient.

IX. Conclusions

The investigation of surface water behavior on an accreting glaze ice surface has shown the following:

1) Surfaces have been observed to have distinct zones of surface water behavior. They include a smooth wet zone in the stagnation region where uniform film runback occurs, a rough zone where surface tension causes coalescence of the surface water into beads and no runback occurs, and a runback zone where surface water runs back as rivulets.

2) The location of the transition point between the smooth and rough zone was observed to propagate with time toward the stagnation point.

3) The freezing of the coalesced water beads in the rough zone generates a characteristic rough surface that enhances heat transfer.

4) Initial surface roughness patterns were observed to propagate in the accreted ice within the "rough zone."

5) Large variations in ice accretions were observed when the surface tension of the water was changed. This illustrates the importance of surface water behavior.

6) Measurement of contact angle and contact angle hysteresis for water on an ice surface showed a strong variation with the ice surface temperature. This temperature dependence indicates that thermal gradients along the ice surface may be important to surface water behavior and to the generation of roughness on an accreting ice surface.

7) Based on this investigation, a simple multiple-zone modification to the current glaze ice accretion model is proposed. The model incorporates discrete zones of surface water behavior each with a characteristic surface roughness.

Acknowledgments

This work was supported in part by NASA and the Federal Aviation Administration under Grants NAG-3-666 and NGL-22-009-640. The work was also supported by the National Science Foundation Presidential Young Investigators Program, Award No. 8552702. Use of the icing wind-tunnel facility was provided courtesy of Data Products of New England, and photographic data were provided by Dr. William Olsen of the NASA Lewis Research Center.

References

[1]MacArthur, C. D., "Numerical Simulation of Airfoil Ice Accretion," AIAA Paper 83-0112, Jan, 1983.

[2]Achenbach, E., "The Effect of Surface Roughness on the Heat Transfer from a Circular Cylinder to the Cross Flow of Air," *International Journal of Heat and Mass Transfer*, Vol. 20, April 1977, pp. 359–369.

[3]Olsen, W. and Walker, E., "Closeup Motion Pictures of the Icing Process," NASA Lewis Research Center Film, 1983.

[4]Olsen, W. A. and Walker, E., "Experimental Evidence for Modifying the Current Physical Model for Ice Accretion on Aircraft Structures," NASA TM 87184, May 1986.

[5]Olsen, W., Shaw, R., and Newton, J., "Ice Shapes and the Resulting Drag Increase for a NACA 0012 Airfoil," NASA TM 83556, Jan. 1984.

[6]Messinger, B. L., "Equilibrium Temperature of an Unheated Icing Surface as a Function of Airspeed," *Journal of the Aeronautical Sciences*, Vol. 20, Jan. 1953, pp. 24–42.

[7]Ludlam, F. H., "The Heat Economy of a Rimed Cylinder," *Quarterly Journal of the Royal Meteorological Society*, Vol. 77, Oct. 1951, pp. 663–666.

[8]Macklin, W. C. and Payne, G. S., "A Theoretical Study of the Ice Accretion Process," *Quarterly Journal of the Royal Meteorological Society*, Vol. 93, April 1967, pp. 195–213.

[9]Gent, R. W. and Cansdale, J. T., "The Development of Mathematical Modeling Techniques for Helicopter Rotor Icing," AIAA Paper 85-0336, Jan. 1985.

[10]Ruff, G. A., "Development of an Analytical Ice Accretion Prediction Method (LEWICE)," Sverdrup Technology Inc., NASA Lewis Research Center Group Progress Report, Feb. 1986.

[11]Dussan, V. E. B. and Tao-Ping Chow R., "On the Ability of Drops or Bubbles to Stick to Non-horizontal Surfaces of Solids," *Journal of Fluid Mechanics*, Vol. 137, Jan. 1983, pp. 1–29.

[12]Hobbs, P. V., *Ice Physics*, Oxford Univ. Press, London, 1974.

THE INFLUENCE OF ICE ACCRETION PHYSICS ON THE FORECASTING OF AIRCRAFT ICING CONDITIONS * +

R. John Hansman, Jr.
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology

ABSTRACT

The physics which control aircraft ice accretion are reviewed in the context of identifying and forecasting hazardous icing conditions. The severity of aircraft icing is found to be extremely sensitive to temperature, liquid water content and droplet size distribution particularly near the transition between rime and mixed icing. The difficulty in measurement and the variability of these factors with altitude, position and time coupled with variable aircraft sensitivity make forecasting and identifying icing conditions difficult. Automated Pilot Reports (PIREPS) are suggested as one mechanism for improving the data base necessary to forecast icing conditions.

1. INTRODUCTION

The accurate and precise forecasting of meteorological conditions favorable to aircraft ice accretion is difficult for several reasons. First, the type and severity of ice accretion is often strongly or nonlinearity dependent on environmental parameters such as temperature, liquid water content, cloud droplet size distribution, turbulence level and water phase. Secondly, several of these parameters are difficult to measure or estimate in the forecasting environment. Finally the severity of the ice accretion and its influence on aircraft performance will depend on both the type and flight condition of the aircraft.

This paper will examine those meteorological and aircraft factors which influence the physics of ice accretion in order to gain some insight into the limitations of current forecasting techniques and the requirements for improved icing forecasts.

2. OVERVIEW OF THE ICE ACCRETION PROCESS

The ice accretion process is controlled by two physically distinct subprocesses. The first is the inertial transport of liquid water (either in the form of cloud droplets, rain drops or mixed phase hydrometeors) from the ambient environment to the aircraft surface. Once droplets have impacted the aircraft surface, their freezing becomes controlled by thermodynamic processes. If the heat transfer from the surface is sufficient to remove all the latent heat of freezing of the impinging water than the droplets will freeze on impact resulting in a dry ice surface. The ice shape typically protrudes forward into the airstream and is commonly referred to as rime ice (see Figure 1).

When the heat transfer from the surface is inadequate to remove all of the latent heat from the impinging droplets the ice surface becomes wet. This type of accretion is commonly characterized as glaze ice. In some cases, both wet and dry ice growth can occur at different places on the same body. This situation is referred to as mixed ice growth. Often in glaze or mixed conditions, the resulting ice shape displays two pronounced growth peaks on either side of the stagnation line (see Figure 1). The most severe aircraft performance degradation is typically associated with such horned ice formations (Renaudo, et al., 1984). Because the physics of mixed and glaze icing are similar, the term mixed icing will refer to both mixed and glaze unless otherwise noted.

The physical processes which control ice accretion are distinctly different for dry and wet ice growth. In dry growth, where the droplets freeze on impact, the ice accretion is controlled by the local rate of impingement of liquid water on the surface. The local impinging mass flux is an inertially determined quantity which involves the individual droplet trajectories as they pass through the flowfield surrounding the body. For wet growth, the ice accretion is controlled by the rate at which latent heat of fusion can be removed from the surface. The heat transfer behavior of the ice surface, therefore, becomes the controlling mechanism for wet ice growth. For mixed icing conditions both the impingement and heat transfer mechanisms play important roles in the ice accretion process.

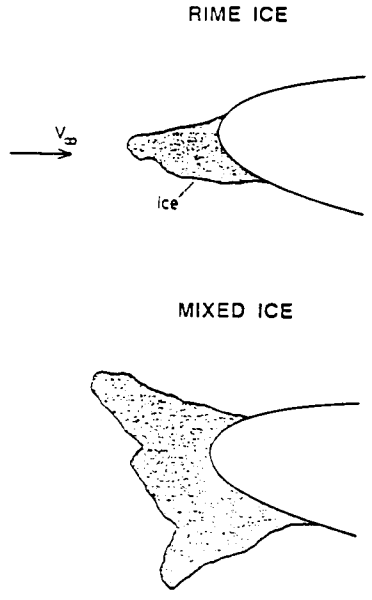


Fig. 1 Typical dry (rime) and wet (mixed/glaze) ice shapes.

2.1 Water Impingement

The physics which control the impingement of liquid water onto the aircraft surface are fairly straightforward. Because the surface is not permeable to gas, flow streamlines do not intersect the body as can be seen in Figure 2. Water droplets have a higher ratio of inertial to hydrodynamic forces than gas molecules and will tend to cross the streamlines resulting in impingement as shown in Figure 2.

* Preprint, Third International Conference on the Aviation Weather System, 1989.
+Some original figures were not available at time of publication.

Much work has been done on droplet impingement trajectories to determine local collection efficiencies β (Bergrun, 1947, Bragg, et al.,1981, Brun, et al, 1953, Gelder, et al, 1956, Hansman, 1984)., The local collection efficiency is defined as the fraction of the mass flux locally impinging onto the surface to the freestream mass flux. The collection efficiency is typically highest near the stagnation point of the body and decreases downstream. The point at which β becomes zero is defined as the impingement limit. The collection efficiency is a strong function of droplet size and body geometry. Large droplets have high inertia tending to cross streamlines resulting in high collection efficiencies and impingement limits. Small droplets tend to follow the streamlines resulting in lower collection values. Small bodies are more efficient droplet collectors because there is less room for the droplets to turn prior to impact.

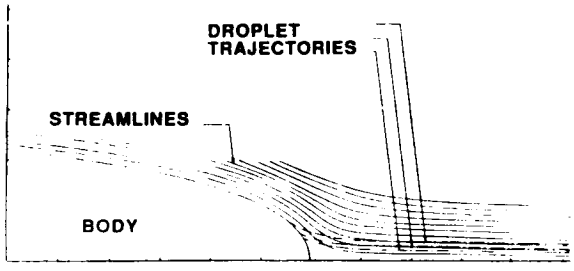


Fig. 2 Example of the relation of water droplet trajectories to the streamline field from Bergrun (1947).

2.2 Thermodynamic Heat Balance

As described above, the thermodynamic heat balance on the accreting ice surface is a controlling factor in determining the rate of ice accretion in mixed icing conditions and is the critical factor in determining the transition between these conditions and rime ice growth. Figure 3 shows the principal modes of energy transfer associated with an icing surface, as depicted by Messinger (1953). Heat is added to the surface primarily from the latent heat of fusion released as the droplets freeze, but also from aerodynamic heating and, to an even smaller extent, from the kinetic energy of the droplets impacting the surface. Heat is removed from the surface primarily by convection, and to a lesser degree by sublimation (when the surface is dry) or evaporation (when the surface is wet). In addition, heat is absorbed from the surface as the supercooled droplets impinge and warm to 0° C. The parameters which primarily influence the heat balance are the temperature difference between the surface and the free stream, the convective heat transfer and the impinging liquid water mass.

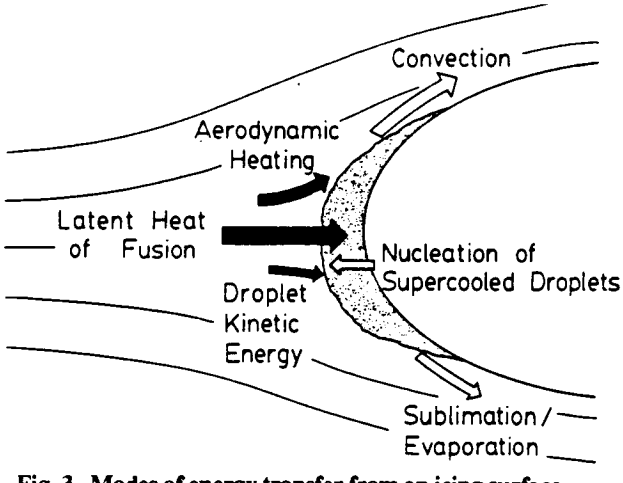


Fig. 3 Modes of energy transfer from an icing surface.

3.0 INFLUENCE OF METEOROLOGICAL PARAMETERS

The influence of various meteorological parameters which are considered important aircraft icing are discussed below in terms of their effect on the physics and the severity of the ice accretion.

3.1 Temperature

Air temperature is one of the most important of the icing parameters. Meteorologists normally work with the ambient or Outside Air Temperature (OAT) however pilots and aircraft designers also use Total Air Temperature (TAT) to include aircraft velocity effects. The TAT is the temperature at the stagnation point of the aircraft and corresponds to the OAT plus an additional temperature rise due to the deceleration of the incoming flow. This "ram rise" can be significant at high velocities. In Figure 4 it can be seen that the difference between stagnation and ambient temperature (TAT-OAT) may be as high as 30° at 500 kts. Because the surface temperature can be lower than the TAT aft of the stagnation point of the aircraft the normal procedure in jet aircraft is to operate anti-icing equipment at TAT values between +10° C and -10° C in the presence of visible moisture.

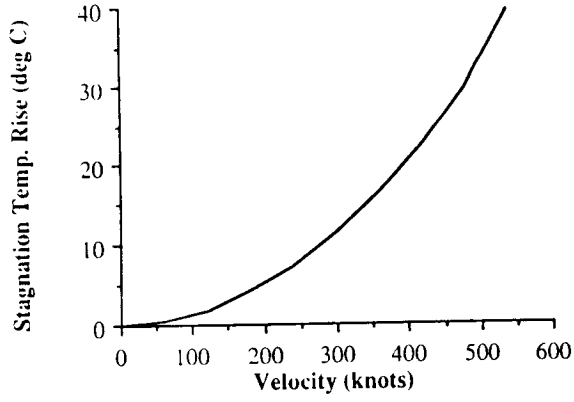


Fig. 4 Stagnation temperature rise versus velocity.

An example of the effect of temperature on ice accretion and performance degradation can be seen in Figure 5. which shows ice shapes and section drag coefficients for a NACA 0012 airfoil at 4° angle of attack obtained by Olsen, Shaw (1984) in the NASA Lewis Icing Research Tunnel. At cold temperatures, the ice accretion was insensitive to temperature and rime accretions were observed with drag increases of 2 to 3 times the clean values. However as the temperature increased above -10° C the drag increased sharply with temperature to a peak value of over 8 times the clean drag. What happened was that the ice growth transitioned from a dry rime growth to a wet mixed growth. The horns characteristic of mixed growth can be observed in the high drag ice accretions. As the Total Air Temperature nears 0° C the ice accretions and the resulting performance degradations decrease due to insufficient heat transfer to freeze all of the incoming water.

From this example it is clear that a relatively small temperature change can cause transition from a relatively benign rime icing condition to a dangerous mixed condition. The nonlinear dependance of icing severity on Total Air Temperature (TAT) combined with aircraft velocity effects make it difficult to accurately identify regions of moderate or severe ice potential from the ambient air temperature alone.

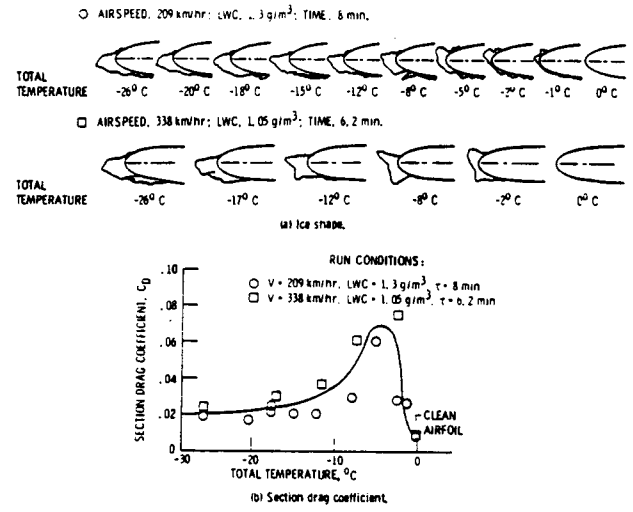


Fig. 5 Effect of total temperature on ice shape and drag. MVD = 20μm; 0.053 m cloud NACA 0012 airfoil at 4° angle of attack. Taken from Olsen, Shaw and Newton (1984).

3.2 Liquid Water Content

Liquid Water Content (LWC) influences the severity of the icing in two primary ways. First, increasing LWC implies more potential water and larger accretions within a given time. Thus, high LWC implies a greater urgency and therefore severity of the icing encounter. The second effect of high LWC is to cause the ice accretion to transition from rime to mixed icing due to the higher impinging water load on the icing surface. This can be seen in the schematic plot of icing severity versus LWC in Figure 6. Rime ice growth will occur even at relatively warm supercooled temperatures for low liquid water contents. In this regime there is a linear increase in the icing severity with LWC. At some value of LWC however, the growth will transition from rime to mixed and the severity of the icing will increase. At colder temperatures the threshold for the transition will occur at a higher liquid water content.

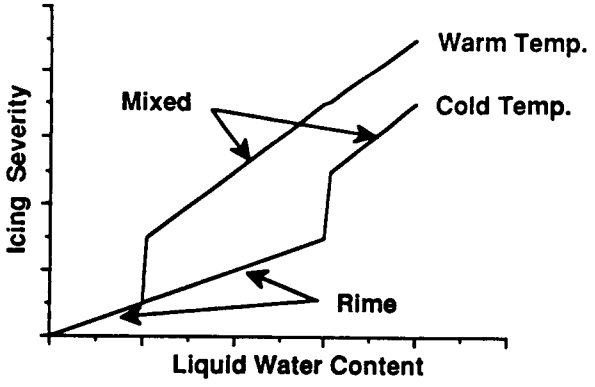


Fig. 6 Schematic plot of icing severity versus liquid water content.

An additional factor which complicates the the ice accretion process is the large variability in LWC often within the same cloud. Figure 7 shows an example of LWC measured by Hansman and Kirby (1987) with a Johnson-Williams hot-wire probe mounted in the nose of the NASA "Twin Otter" Icing Research Aircraft. During this flight, the

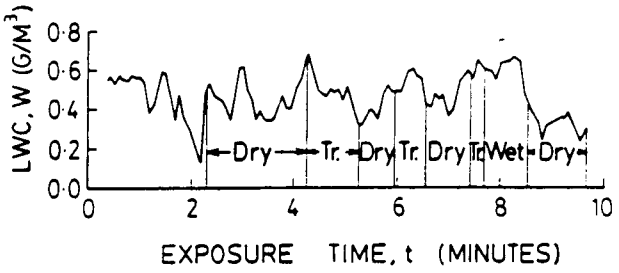


Fig. 7 Plot of liquid water content versus exposure time showing typical fluctuations observed in natural icing conditions. Also shown are ultrasonically measured periods of wet, dry and transitional ice growth.

stagnation point ice accretion was monitored by ultrasonic techniques which could determine weather the accretion was dry, wet or transitional. It can be seen that the variation in liquid water content caused the accretion to vary from dry growth characteristic of rime accretion to wet growth characteristic of mixed accretion within the same cloud.

3.3 Droplet Size

The size of the ambient water droplets both in terms of the Median Volumetric Diameter (MVD) and the actual shape of the Droplet Size Distribution can be important to the ice accretion process. As described in Section 2.1, large droplets are more efficiently caught by body. These effects have been quantified for typical cloud and rain size distributions by Hansman (1984). Figure 8 shows the impinging mass distribution function and the ambient size distribution function for a Khrigian-Mazan distribution with 20 micron mean effective droplet diameter. It can be seen that the bulk of the impinging mass results from the small number of large droplets in the tail of the distribution. The MVD and the shape of the distribution therefore determine the effective collection efficiency of the aircraft. It is not uncommon to have trace or negligible icing even at high liquid water contents for clouds of predominantly small droplets. It is also possible to have moderate to severe icing at relatively low liquid water contents for distributions consisting primarily of large droplets.

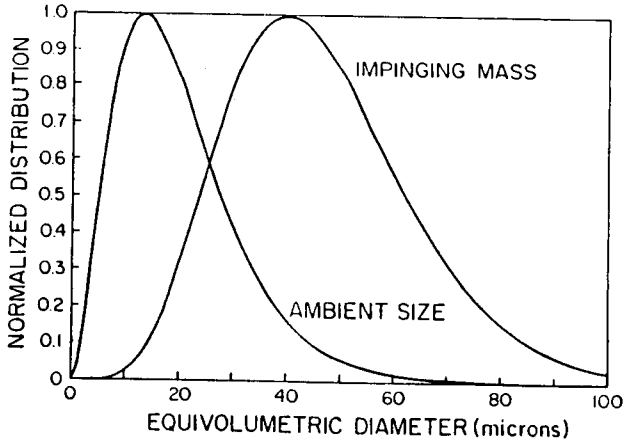


Fig. 8 Khrigian-Mazan ambient cloud droplet size distribution and the resulting impinging mass flux distribution.

The presence of large droplets within the cloud distribution can result in additional hazard due to increased impingement limits. The limit of droplet impingement on an

aircraft component increases with droplet size. Current design guidance (FAA AC-20-73, 1971) recommends that a diameter of 40 microns be used to determine impingement limits. The presence of significant numbers of droplets in excess of 40 microns can therefore result in ice accretions occurring behind the protected regions of the aircraft. These factors are thought to contribute to the anomalously high performance degradations observed by Cooper, Sand, Politovich and Neal (1984) in clouds with droplets ranging from 40 to 300 microns.

The problems resulting from large droplets are exacerbated in freezing rain where both large droplet sizes and large liquid water contents are combined. Freezing rain commonly results in clear glaze ice accretions with significant runback icing. It is considered an extremely hazardous condition.

3.4 Cloud Phase

The icing potential for a particular cloud is directly related to the phase of the hydrometeors. As described above, icing normally results from the impact of supercooled water droplets. Dry ice crystals, generally, do not adhere to the aircraft surfaces after impact and are therefore not considered an icing hazard. If however, the ice crystals are wet due to partial melting or the aircraft surface is wet due to de-icing or the recent penetration of a high LWC region then the impinging crystals will stick.

Mixed phase icing (not to be confused with mixed rime/glaze icing) is relatively rare. Some examples of mixed phase icing have however, been observed in flight tests by Gayet, Bain and Soulage (1984) who noted that the presence of snow in supercooled clouds significantly reduced the icing rate. Ice crystal sticking on wet aircraft surfaces was observed during NASA Icing Research Aircraft flight tests of ultrasonic ice detector arrays where the wet aircraft surface was documented by ultrasonic techniques (Hansman, Kirby, McKnight and Humes; 1988).

In general, forecasting efforts are directed towards identifying regions of supercooled cloud. Techniques are available to predict cloud phase in stratiform cloud and glaciation results in overestimation of the icing severity (USAF, 1980). However, in cumuliform clouds, cloud phase uncertainty represents a potential source of error in the forecasting of icing conditions.

3.5 Fine Scale Turbulence Level

An additional factor which has recently emerged as potentially important to the ice accretion process is the fine scale (centimeter and below) ambient turbulence level. The turbulence level is known to strongly influence the convective heat transfer from the icing surface. As shown in Section 2.2, the convective heat transfer is one of the primary parameters in determining the transition between rime and mixed icing with the resulting effect on icing severity. Flight test observations by Hansman and Kirby (1987) observed a wide variability in the parametric threshold between rime and mixed icing. The variability was thought to be due to variations in heat transfer resulting from the ambient turbulence level. While the effect of fine scale turbulence level on icing severity has not been directly demonstrated it may be an additional source of uncertainty in the forecasting process.

4.0 INFLUENCE OF AIRCRAFT PARAMETERS

The influence of various aircraft parameters which are considered important to the icing problem are discussed below in terms of their effect on the physics and severity of the ice accretion.

4.1 Velocity

The aircraft velocity effects both the collection of liquid water and the thermodynamics of the icing process. Increasing velocity results in higher impinging liquid water exposure by increasing both the path swept out by the aircraft trajectory in a give time and the collection efficiency.of the aircraft surfaces. Thermodynamically, the velocity effects the heat load through the increased impinging water mass and through increasing the stagnation point temperature at high velocities. This is shown in Figure 4 where the stagnation temperature rise is plotted as a function of velocity.

4.2 Shape

The shape of the accreting body has a large effect on the local collection efficiency. Generally, smaller bodies are more efficient collectors than larger bodies. Therefore, slender components such as propellers, fan blades and antennas will tend to be the most sensitive to ice accretion. As a result of their high collection efficiencies, windshield wipers or Outside Air Temperature probes are often used by crews to detect icing conditions in flight.

Three dimensional shape effects can also be important for many aircraft components. For example wing sweep commonly results in a spanwise variation of the accretion in mixed icing conditions. This "lobster tail" ice can result in significant performance degradation. It should be noted that current forecasting procedures are based on straight wing propeller driven aircraft (USAF, 1980).

4.3 Aircraft Category

The effect of icing varies significantly with individual aircraft design. While it is beyond the scope of this paper to discuss the icing sensitivity of individual aircraft, the sensitivity of broad aircraft categories will be discussed briefly below.

4.3.1 Turbojet/Turbofan Aircraft

Jet aircraft are considered to be the least susceptible to icing. Jet aircraft normally operate with significant quantities of excess thrust which can be used to offset performance degradation. In addition, the typical flight profile of a jet aircraft is to climb and descend rapidly through the lower troposphere where the icing potential is greatest and to cruise at high altitudes (20,000 ft to 45,000 ft.). Occasionally, Air Traffic Control (ATC) requirements will dictate sustained operation at low altitudes particularly in busy terminal areas. The primary icing hazard to jet aircraft is engine failure due to Foreign Object Damage (FOD). This results from the ingestion of chunks of ice which are shed off of other aircraft components such as engine inlets or in some aircrafts the wings. Therefore, critical regions are normally anti-iced with hot bleed air from the engines to prevent any ice accumulation

4.3.2 Turboprop and Reciprocating Aircraft

There is a tremendous variability in the sensitivity of propeller driven aircraft to icing. Large turboprops tend to be fully ice protected and can operate successfully in regions of high icing potential. Small reciprocating engine aircraft are generally not equipped with ice protection and are not approved for flight in icing conditions. Even light icing conditions are a potential hazard to unprotected aircraft and forecasting uncertainties have the greatest impact on this aircraft category.

Propeller driven aircraft operate at low altitudes where icing potential may exist over the entire flight. The most critical components are generally the propellers because

loss of propeller efficiency translates directly into loss of thrust. However, even airframe icing can pose a significant hazard because propeller driven aircraft normally operate at much lower excess thrust margins than jet aircraft. Other hazardous factors include; reduced stability, loss of control authority and reduced visibility due to windshield icing.

4.3.3 Rotorcraft

Helicopter icing has become important during the last decade where helicopter flight in Instrument Meteorological Conditions (IMC) has become more commonplace. Helicopter operations occur almost exclusively at low altitudes where there is significant icing potential. Helicopters are extremely susceptible to icing conditions. Rotor icing simultaneously degrades the lift and thrust efficiency of the vehicle. In addition, helicopters typically operate with very slim power margins and can therefore only tolerate minimal ice accretion. Other hazardous factors resulting from helicopter operations in icing conditions include; blockage of engine inlets, reduced control authority, vibration due to asymmetrical ice load on rotors and reduced visibility due to canopy icing.

5.0 IMPLICATIONS FOR THE FORECASTING OF ICING CONDITIONS

Some of the difficulties in forecasting or identifying hazardous icing conditions are clear from the above. Many of the parameters which were shown to be important to the ice accretion process such as the droplet size distribution and cloud phase are not available to the forecaster. Other parameters, which may be available, such as temperature or liquid water content are nonlinearly related to icing severity. Limitations in accuracy and spatial resolution of these parameters result in over or under prediction of the icing severity. Finally the variable susceptibility of different aircraft types implies that a single icing hazard analysis will result in over or under prediction of the severity for certain aircraft categories.

Given the limitations on the forecasting process, current techniques do a remarkably good job at identifying general regions of potential icing conditions. One of the key indicators which are used to validate or initiate an icing forecast are pilot reports (PIREPS). By actually penetrating the icing environment, aircraft can directly measure the severity of the icing condition. One of the difficulties with PIREPS is that the reports are not well calibrated and the variable susceptibility of aircraft discussed in Section 4.3 must be considered in their interpretation. Another problem is the timely dissemination and generation of PIREPS. Because of other responsibilities, ATC cannot always process PIREPS rapidly which may discourage the voluntary pilot reports.

One potential improvement, which has been implemented to a limited extent,is the automated generation and transmission of PIREPS over digital data links. By the use of onboard ice accretion sensors, automated continuous reporting of the presence or lack of icing conditions could be accomplished. If an adequate fleet of aircraft were so equipped, a significant improvement in the forecasting and identification of hazardous icing conditions could be realized. Other potential technical developments which may improve icing forecasting include: vertical sounders, improved satellite imaging, and improved data synthesis systems such as the PROFS system.

6.0 CONCLUSIONS

The severity of aircraft icing is found to be extremely sensitive to temperature, liquid water content and droplet size distribution particularly near regions of transition between rime and mixed icing conditions. The difficulty in measurement and the variability of these factors with altitude, position and time coupled with variable aircraft sensitivity make forecasting and identifying hazardous icing conditions difficult. Automated Pilot Reports (PIREPS) are suggested as one mechanism for improving the data base necessary to forecast icing conditions.

ACKNOWLEDGMENTS

This work was supported by the National Aeronautics and Space Administration, the Federal Aviation Administration under the Joint University Program for Air Transportation, Grants NGL-22-009-640 and NAG-3-666, and the National Science Foundation Presidential Young Investigator Award Program.

REFERENCES

Bergrun, N.R., 1947: A Method for Numerically Calculating the Area and Distribution of Water Impingement on the Leading Edge of an Airfoil in a Cloud. NACA TN-1397.

Bragg, M. B., Gregorek, G. M., and Shaw, R. J., 1981 : Analytical Approach to Airfoil Icing. AIAA Paper 81-0403.

Brun, R. J., and Mergler, H. W., 1953 : Impingement of Water Droplets on a cylinder in an Incompressible Flow Field and Evaluation of Rotating Multicylinder Method for Measurement of Droplet-Size Distribution, Volume-Median Droplet Size, and Liquid Water-Content in Clouds. NACA TN2904.

Cooper, W. A., Sand, W. R., Politovitch, M. K., Veal, D. L., 1984 : Effects of Icing on Performance of a Research Airplane. *Journal of Aircraft*, 21, pp. 708-715.

FAA, 1971: Advisory Circular. AC-20-73

Gayet, J. F., Bain, M., and Soulage, R. G., 1984: Role of Ice Crystals on Ice Accretion Processes. *Proceedings of the Second Int'l Workshop on Atmospheric Icing of Structures*, Norway.

Gelder, T. F., Smyers, W. H., Jr., and von Clahn, U., 1956 : Experimental Droplet Impingement on Several Two-Dimensional Airfoils with Thickness Ratios of 6 to 16 Percent. NACA TN3839.

Hansman. R. J., Jr., 1984 : The Effect of the Atmospheric Droplet Size Distribution on Aircraft Ice Accretion. *Journal of Aircraft*, 22, pp. 503-508.

Hansman, R. J., Jr., and Kirby, M. S.,1987: Comparison of Wet and Dry Growth in Artificial and Flight Icing Conditions. *Journal of Thermophysics and Heat Transfer*, 1, pp. 215-221..

Hansman, R. J., Jr., Kirby, M. S., McKnight R. C., and Humes, R. L.,1988: In-Flight Measurement of Airfoil Icing Using an Array of Ultrasonic Transducers. *Journal of Aircraft*, 25, pp. 531-537.

Messinger, B. L., 1953 : Equilibrium Temperature of an Unheated Icing Surface as a Function of Airspeed. *Journal of the Aeronautical Sciences*, pp. 24-42.

Olsen, W., Shaw, R., and Newton, J., 1984 : Ice Shapes and the Resulting Drag Increase for a NACA 0012 Airfoil. NASA TM-83556.

Renaudo, R.J., Mikkelson, K. L., McKnight, R.C., and Perkins, P. J., Jr., 1984 : Performance Degradation of a Typical Twin Engine Commuter Type Aircraft in Measured Natural Icing Conditions. NASA TM-83564.

USAF Air Weather Service, 1980: Forecaster's Guide to Aircraft Icing. AWS/TR-80/001

Cockpit Display of Hazardous Weather Information*#

R. John Hansman, Jr.* and Craig Wanket
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, Massachusetts

Abstract

Information transfer and display issues associated with the dissemination of hazardous weather warnings are studied in the context of windshear alerts. Operational and developmental windshear detection systems are briefly reviewed. The July 11, 1988 microburst events observed as part of the Denver TDWR operational evaluation are analyzed in terms of information transfer and the effectiveness of the microburst alerts. Information transfer, message content and display issues associated with microburst alerts generated from ground based sources (Doppler Radars, LLWAS and PIREPS) ars evaluated by means of pilot opinion surveys and part task simulator studies.

1. Introduction

Technological advances in ground-to-cockpit datalink capability, information display, and hazardous weather detection create the possibility for new and improved methods of informing flight crews about weather hazards. However, the availability of increased information and multiple modes of communication also lead to problems of system integration. Issues including the selection, transfer, and presentation of information must be addressed in the development of advanced systems for the display of hazardous weather information. In addition, design procedures, centered around the needs of the flight crew and the capabilities of the available equipment should be applied.

The display and information transfer issues related to advanced windshear alerting systems in the terminal area have been chosen as an initial point of focus. This problem was chosen both to investigate general issues related to the dissemination of hazardous weather information and to focus on specific issues of a critical near term need. Windshear in the terminal area is one of the most dangerous weather-related problems faced by aviation today.[1]. The real-time detection of windshear hazards is a very active field of research,[2,3,4] and thus provides a useful testing ground for issues related to advanced data uplink and display of hazardous weather information.

2. Background

2.1 Terminal Area Windshear

Low-altitude windshear is the leading weather-related cause of fatal aviation accidents in the U.S. Since 1964, there have been 26 accidents attributed to windshear resulting in over 500 fatalities [1,5]. Low-altitude windshear can take several forms. Macroscopic forms, such as gustfronts caused by colliding warm and cold air masses, can generally be predicted and avoided. However, the small intense downdrafts known as microbursts are far more dangerous and difficult to detect.

Microbursts begin with a cool downdraft formed at the base of a cumulus or cumulonimbus cloud. If the downdraft is strong enough to impact the surface, it spreads out radially and creates a small area (1 to 4 km in diameter) of intense windshear. Such conditions typically last for short periods (10-30 min), but can be very dangerous to aircraft at low altitudes, particularly on takeoff or final approach Initially, the aircraft experiences a strong headwind, which causes a momentary increase in lift. Next, the aircraft enters an area of downdraft, and then a sharp tailwind. This combination results in loss of effective airspeed and corresponding loss of lift. (Fig. 1). It may also serve to destablilize the flight trajectory. The resulting performance loss can in some cases be sufficient to result in ground impact. In addition, microbursts can be accompanied by strong edge vortices, which can further destabilize the aircraft. Most fatal windshear accidents have been attributed to microbursts.[5]

An additional factor which makes microbursts particularly dangerous is that they are generally not obvious either visually or to standard airborne weather radar. Microbursts have been observed to occur both during periods of severe rain or during periods of little or no low-altitude precipitation. For meteorological and instrumentation purposes, it is convenient to distinguish between 'wet' and 'dry' microbursts. Dry microbursts, more common in the western U.S., can sometimes be detected by the presence of curling clouds of dust on the ground or vertical cloud shafts known as 'virga'. Wet microbursts cannot generally be distiguished from benign rain cells with radar reflectivity information.

Microbursts have been observed with intensities greater than most aircraft could be reasonably expected to survive. Avoidance is the best way to handle a windshear hazard. This indicates a need for reliable remote detection, allowing the flight crew adequate advance warning to plan and execute a maneuver to avoid microburst penetration.

Fig. 1: Microburst windshear encounter on approach (from Ref. 2).

*Associate Professor, Associate Fellow AIAA
†Research Assistant, StudentMember AIAA

*AIAA-89-0808. #Some original figures were not available at time of publication.

2.2 Microburst Detection

2.2.1 Current Procedures

Current procedures for microburst detection and warning center around the Low-Level Windshear Alert System (LLWAS), Pilot reports, and improved pilot education through efforts such as the FAA's *Windshear Training Aid*.[6] LLWAS is a system of anemometers currently in service at most major U.S. airports designed to measure shifts in wind speed and direction within the airport perimeter. Although capable of detecting macroscopic phenomena such as gustfronts, the anemometer spacing is larger than the characteristic surface dimension of many microbursts, and thus LLWAS remains fairly ineffective for detection of microburst windshear. The *Windshear Training Aid* states that "If an LLWAS alert (triggered by wind speed and/or direction differential) occurs, it indicates the presence of something shear-like, though not necessarily indicative of magnitude or location. However, the absence of an alert does not necessarily indicate that it is safe to proceed!".[6]

Pilot reports (PIREPS) of windshear provide the most reliable data. The availability of PIREPS necessarily requires that an aircraft penetrate a microburst, which is not desirable; but the information, unlike LLWAS, provides conclusive evidence of a windshear hazard for subsequent aircraft. It is desirable to integrate PIREPS with any sensor data available in future windshear detection systems.

The *Windshear Training Aid* itself is designed to inform pilots and controllers about windshear, primarily how to recognize and avoid or recover from microburst encounters. Avoidance is practiced through the use of LLWAS information, weather reports, and visual clues. In past accidents, these clues have been largely ignored; increased windshear training emphasis is being used to increase pilot awareness of these events. However, high pilot workload in the terminal area and the relative rarity of hazardous windshear makes it difficult for crews to fully assimilate the evidence of windshear before penetration.

2.2.2 Emerging Windshear Detection Technologies

To meet the need for improved windshear warning, new systems for detection are under development. Both airborne and ground-based systems are under consideration. Airborne look-ahead systems are still primarily experimental: candidate technologies include doppler radar, doppler lidar, and infrared radiometry. [2,3,4] To be an effective, dependable windshear avoidance tool, an airborne system must be able to detect windshear ahead of the aircraft to a range of 1 - 3 km, thus typically providing 15 to 45 seconds of warning. Also, the sensor should work for either wet or dry microbursts with enough resolution to adequately measure size and intensity. None of the methods mentioned have yet fully demonstrated these capabilities in flight.

Ground-based remote sensing technology is much more developed. LLWAS and PIREPS often yield useful data, but are not always available or accurate. Ground-based doppler radars have been succesfully demonstrated for microburst detection (JAWS, Huntsville, Denver) [5] and have an advantage over airborne systems in terms of ground clutter suppression, size and power. Experiments performed at Huntsville, AL in 1986 and at Denver in 1987 and 1988 have shown impressive results (Table 1). The predominance of wet microbursts at Huntsville and dry microbursts in Denver shows the versatility of the ground-based doppler radar. The ability of such systems to integrate data aloft with wind measurements near the surface allows for earlier forecasting of microburst locations and outflow strengths.

MICROBURST DETECTION

Data	Probability of detection* $\Delta V<20$ m/s	$\Delta V \geq 20$ m/s	Total	Probability of false alarm
Huntsville 1986	88%	100%	91%	5%
Denver 1987	90%	99%	92%	5%
Combined	90%	100%	92%	5%

GUST-FRONT DETECTION

Data	Probability of detection $\Delta V \leq 15$ m/s	$\Delta V > 15$ m/s	Probability of false alarm
Denver 1987	81%	93%	5%

* ΔV = net wind change in shear region (only events with ΔV values greater than 10 m/s are scored.)

Table 1:Doppler radar windshear detection results [7]

The demonstrated capability of ground-based doppler radar for windshear detection and forecasting makes it the most viable system for near term use for microburst avoidance information. The combination of doppler radar, improved LLWAS systems and PIREPS makes an integrated ground-based system the primary focus for system integration and automated datalink issues.

2.3 Ground-to-Air Data Transfer

Digital ground-to-air data transfer is an area under active development. Several methods of digital ground-to-air data transmission are currently or nearly available. ACARS, a privately-sponsored system for the uplink and downlink of digital information related to commercial aviation, is currently in use by many major airlines. It provides a high-speed alphanumeric datalink for flight management information, helping to relieve congestion on crowded ATC voice frequencies. With the addition of satellite relays, ACARS coverage will extend to most international commercial air routes. The first satellite transmissions are expected to begin in the third quarter of 1989 in the Pacific Ocean region [8].

Another system slated for near-term deployment is the FAA's Mode-S surveillance datalink. Mode-S is an extension of the altitude encoding Mode-C transponder in the ATC Radar Beacon System allowing message delivery from ATC to individual aircraft. Each individual message can carry 48 useful bits of information, and the time for the interrogation beam to scan the entire coverage area is 4 to 12 seconds. Messages can be also be linked in groups of up to 4 frames or sent as a longer Extended Length Message with less urgency.

In the long term, the Aviation Satellite Communications System (SatCom) is being developed. The goal is a standardized worldwide system for digital voice and data communications, based on nine existing satellites in geosynchronous orbit.[9] Other systems such as digital ATIS or enroute weather channels are also envisioned for future development.

2.4 Information Transfer Issues in the 1988 Denver TDWR Evaluation

An event which illustrates many of the information transfer issues occurred during the 1988 Terminal Doppler Weather Radar (TDWR) operational demonstration at Stapelton International Airport in Denver. On July 11, a period of severe microburst activity occurred. It is instructive to evaluate the warnings and responses of the five aircraft which initiated and abandoned approaches immediately prior to the closure of the airport.

2.4.1 TDWR Setup

In the 1988 operational evaluation, the TDWR radar was operated by the MIT Lincoln Laboratory and located at the Buckley AFB southeast of the Stapelton airport. Microburst and gustfront alerts were generated from the doppler weather radar data by an automatic algorithim and confirmed on line by an NCAR meterologist. The alerts were then sent by ground line to the Stapelton control tower and the terminal radar approach control (TRACON).

The information was displayed in two formats in the control tower. The local tower controller, who had primary responsibility for the dissemination of microburst alerts, had an alphanumeric display (Fig. 2) which could present either TDWR or LLWAS information in the same format. This was done to minimize the transition between periods of TDWR and LLWAS-only operation. The tower supervisor and the TRACON also had the geographical situation display which is shown in Fig.3. This color display presented the locations of microbursts, gustfronts and precipitation on a plan view of the runway configuration. In addition, LLWAS wind vectors are displayed. In the tower a local controller working arrivals to Runways 26L and 26R would have to cross the tower cab to have access to the geographical situation display.

Type of wind shear	Runway	Threshold winds	Wind shear Headwind change (kts)	Location
	CF	190 16 G 25		
MBA	35 LD	160 22	50-	RWY
MBA	35 RD	180 5	25-	RWY
MBA	35 LA	030 23	55-	1 MF
	35 RA	180 10	60-	3 MF
MBA	17 LA	180 5	25-	RWY
MBA	17 RA	160 22	55-	RWY
	17 LD	180 10	60-	RWY
MBA	17 RD	030 23	55-	RWY

Fig. 2: Example of controllers alphanumeric display (from Ref. 7).

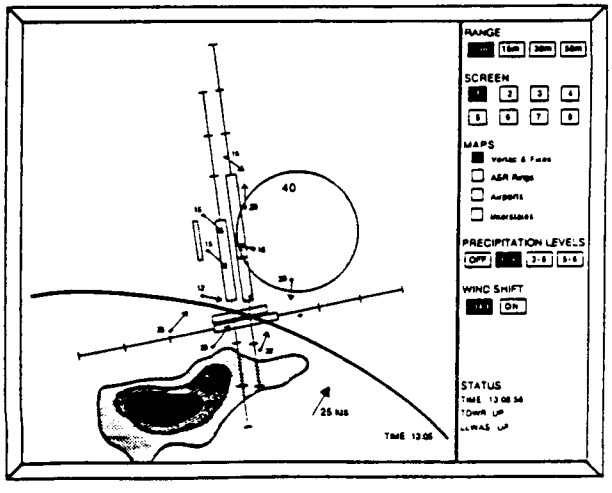


Fig. 3: Geographical situation display used in the control tower and TRACON (from Ref. 7).

2.4.2 July 11, 1988 Scenario

The period of intense microburst activity began at the Stapelton airport shortly after 2200 UTC. At this time arriving aircraft were landing on runways 26L and 26R. Departing aircraft were using runways 35L and 35R. On the arrival ATIS aircraft were informed of a convective SIGMET for the eastern Colorado area,and that the doppler radar windshear detection demonstration was in progress. After 2203 UTC the ATIS was updated to include "low level windshear advisories in effect".

The evolution of the microburst event can be seen in the geographical situation displays presented to the tower supervisor at at 2201, 2207 and 2212 UTC (Fig.4). At 2201 UTC there

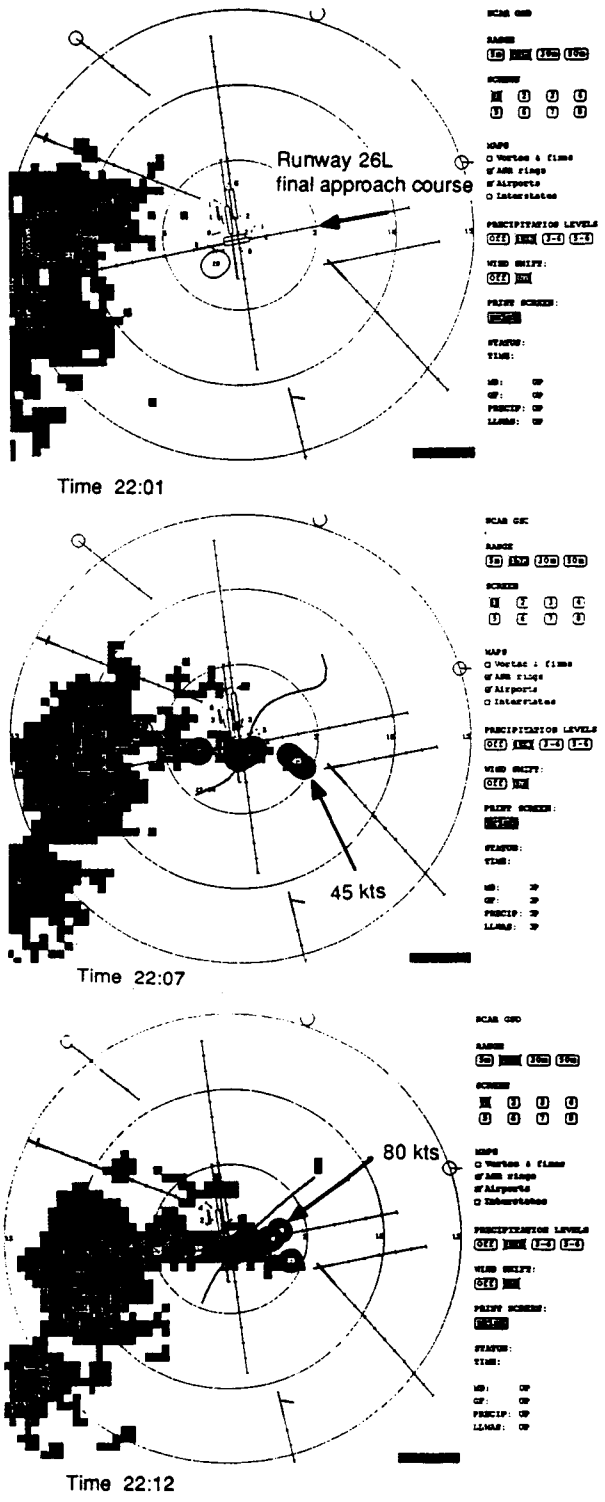


Fig. 4: Geographical situation display plots showing the evolution of the severe microburst event of July 11.

was an area of precipitation southwest of the airport and a region of 25 kt windshear within 2 miles of the airport center. By 2207 UTC. a gustfront had developed over the airport with some light precipitation. Several microbursts had developed with the gustfront including a 45 kt headwind to tailwind cell located on the approach to runways 26L and 26R. By 2212 UTC the microburst had increased in strength to 80 kts and the precipitation had increased. This microburst event continued at high intensity to 2222 UTC when it began to abate. Windshear values of 30 kts were still being measured at 2230.

The altitude versus time plots generated from Mode C transponder replies for the 5 aircraft which initiated approaches between 2207 and 2214 UTC are shown in Fig. 5. Also shown are the times at which microburst alerts were given to the aircraft and the time of reported missed approach. All aircraft which penetrated the microburst reported intense windshear. Transcripts of the verbal microburst alerts given to each aircraft by the local tower controller are presented in Table 2. It is unknown if there were any microburst alerts issued to these aircraft by the TRACON approach controller. However, the fact that 4 of the 5 aircraft elected to continue the approach indicates that this was unlikely.

2.4.3 Implications of the July 11 Experience

Several issues important to the development of microburst alerting systems are apparent from this data. The variability in aircrew interpretation of microburst warnings can be seen by comparing the response of aircraft A to that of aircraft B. The aircraft were approaching parallel runways and were issued virtually identical alerts within 30 seconds of each other. Aircraft A elected to immediately abandon the approach based on the microburst alert and visual observations of a descending rain shaft. This aircraft never penetrated the primary microburst area. Aircraft B elected to continue the approach, penetrated the microburst, and descended to within 100 ft of the runway threshold before executing a missed approach.

Another issue which arises from the data is the delay between the generation and the voice transmission of the alert to the aircraft by ATC. Fig.6 plots the delay to alert for each aircraft based on the first TDWR generated microburst alert at 22:06:17 UTC and the assumption that no alerts were given to these aircraft by the TRACON. It can be seen that the shortest delay was approximately 60 seconds and that a delay of 350 seconds was encountered for the last aircraft to report to the tower (Aircraft E). The delays in excess of 100 seconds are likely a result of the effort to make the TDWR alerts apear like LLWAS alerts. The primary windshear alert responsibility therefore rested with the tower controller who did not have contact with the aircraft until they were at the outer marker. It does appear, however, that a minimum delay of approximately 60 seconds can be expected for the dissemination of verbal alerts even if the aircraft is in contact with the controller who has alerting responsibility.

A third issue which arises is that the initial microburst alert for each aircraft was imbedded within a routine landing clearance message. The routineness of the message may have resulted in a lack of urgency associated with the alert. This possible lack of urgency coupled with the high cockpit workload which occurs at the outer marker may have contributed to the difficulty some crews had in fully assessing the magnitude of the hazard. It is also worth noting that the tower controller relied primarily on the alphanumeric display. It is interesting to consider whether his level of urgency may have increased if he had access to the geographical situation display and could have more easily visualized how the situation was developing.

The final point which comes out of the analysis is the importance of PIREPS. Both the flight crews and the tower controller were more likely to react conservatively to the microburst alert after several aircraft had gone around and

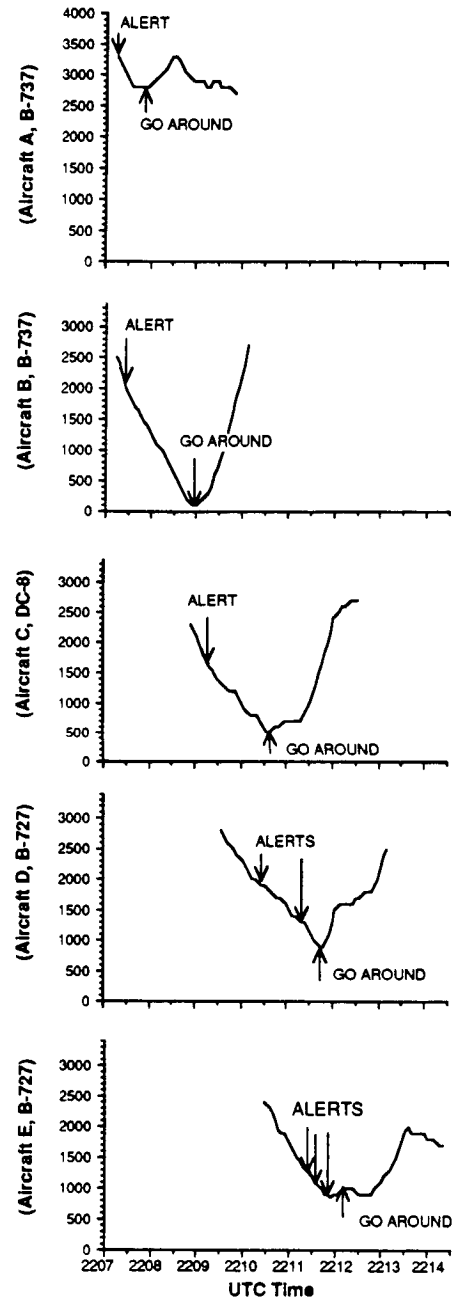


Fig. 5: Plots of aircraft height (AGL) versus time.

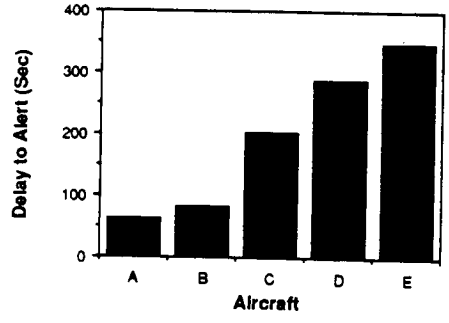


Fig. 6: Delay between first microburst alert and transmission of alert to aircraft.

	Aircraft A
22:07:15	"Aircraft A, Denver tower, runway two six right, cleared to land. Microburst alert, centerfield wind two two zero at none, a forty knot loss, one mile final as reported by machine, no pilot report."
	Aircraft B
22:07:35	"Aircraft B, Denver tower, runway two six left cleared to land. Winds two one zero at five, a forty knot loss, one mile final microburst alert, not substantiated by aircraft."
	Aircraft C
22:09:35	"Aircraft C heavy, Denver tower, microburst alert, threshold wind one four zero at five, expect a fifty knot loss, two mile final, runway two six left, cleared to land."
	Aircraft D
22:11:05	"Aircraft D, caution have turbulence from the heavy DC-8. He is going around. We have a microburst alert, threshold winds, zero nine zero at three. Expect a seventy knot loss on a three mile final."
22:11:45	"Microburst alert, runway two six. Threshold wind, one five zero at five, expect an eighty knot loss on a three mile final."
	Aircraft E
22:12:05	"Aircraft E, microburst alert, threshold wind one six zero at six, expect an eighty knot loss on a three mile final, say request."

Table 2: Transcripts of verbal microburst alerts issued to each aircraft.

reported wind shear. This, coupled with the increasing microburst intensity, explains why the later aircraft initiated their missed approaches at higher altitudes than aircraft B which had no PIREP information to confirm the microburst alert.

3. Research on Windshear Detection and Warning in the Advanced ATC Environment

3.1 Problem Statement

The integration of ground-based information sources with digital datalinks such as Mode-S shows great potential for the accurate prediction and delivery of microburst windshear alerts with minimal delay. Fig. 7 illustrates possible information flow configurations for such a system. The multiple potential data paths are dependent on the acceptable degree of automation. Clearly, the delay between detection and alert is minimized with a fully automated process whereby computer algorithms determine alerts from PIREPS, TDWR, and LLWAS data and use Mode-S to directly distribute them. However, putting the controller in the loop to some degree would help filter false alarms and more efficiently control the destination of the data.

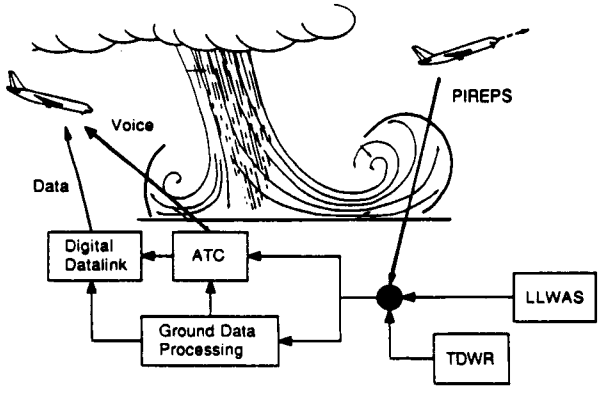


Fig 7: Possible windshear data distribution in the advanced ATC environment.

A number of other information issues also require consideration. The content, timing, transmission, and presentation of windshear information all need to be determined. Automated links such as Mode-S are subject to bit limitations and reliability considerations. This has an impact on message content and distribution. The timing and priority of alerts must be considered to get maximum efficiency and to generate the least possible confusion. The high workload of both controllers and flight crews during terminal area operations adds a further measure of difficulty. Finally, the varying levels of instrument sophistication in civil aircraft must be considered. The advanced moving map displays in modern transport category aircraft allow for development of user-oriented graphical presentations, while many general aviation aircraft have no visual display capability.

3.2 Investigations

Several investigations are being performed to address the issues discussed above. Flight crew opinion surveys are being used to obtain user input on a number of factors. Data is sought on current operational issues such as LLWAS and other available windshear information sources, as well as pilot perceptions of the microburst threat. Also, issues of data transmission and presentation are addressed. In addition, some issues are being addressed through flight simulation studies. A simple experiment based on a general aviation simulator was conducted to compare voice communication with graphical data presentation modes. Also, a part-task simulation of the Boeing 757/767 has been developed in order to do more sophisticated investigation into optimization of graphical warning formats, information content and delivery timing, and the effect on pilot workload.

3.3 Results

3.3.1 Current Windshear Procedures

For user input on current windshear alert systems and requirements for future systems, a pilot opinion survey is being conducted. A preliminary sample of 20 United Airlines line and training pilots has been completed, and a further distribution of 250 is under way. Initial results show several consistent trends. It is almost universally agreed (94%) that microbursts pose a major safety hazard to transport aircraft. Fifty-three percent of the respondents have had what they considered to be a hazardous windshear encounter; most incidents occurred at DEN, a United operations hub. When posed the question "Currently available windshear alert data is sufficient for safe operation in the terminal area," only 17% of the respondents agreed, while 56% disagreed. All but one of the pilots felt that "...a system to provide aircrews with better and more timely windshear alerts is necessary." The results clearly indicate that flight crews are not completely confident in currently available data and would be very receptive to improvements. Figure 8 shows the pilots' average ranking of possible sources of windshear information.

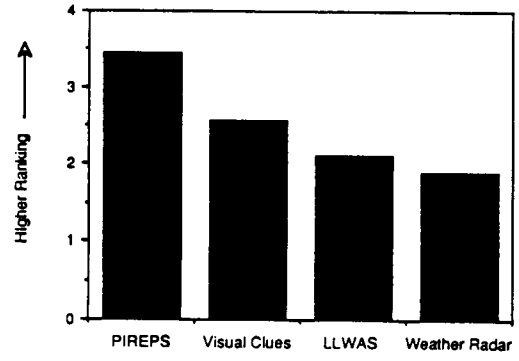


Fig. 8: Pilot ranking of windshear information sources.

Significantly, PIREPS and visual clues are both considered more useful for windshear avoidance than LLWAS alerts. Yet, neither pilot reports or visual clues are always available; this underlines the need for a detection system which can reliably provide some degree of advance warning.

3.3.2 Modes of Information Presentation

Because of the high workload in terminal area operations, it is important to consider the manner in which information is presented to the flight crew. This was illustrated by the Stapleton incident; even though data was available, it was difficult to effectively communicate it to the flight crew. There are several possible modes of information presentation in the cockpit: *voice*, *alphanumeric*, or *graphical*. Issues to be considered include crew workload, preferences, and the capabilities of the aircraft instrumentation. The widespread use of CRT displays in modern transport aircraft, for example, opens up new possibilities for totally automated graphical information displays. Moving map displays, such as the Electronic Flight Instrumentation System (EFIS) used on the Boeing 757-767 generation of aircraft, are good candidates for display of critical weather information.

Responses from the pilot survey indicate that pilots are receptive to graphic displays. (Fig. 9). The specific suggestion of integrating windshear information with an EFIS-type moving map display was strongly supported. Also of interest was the preference of ATC voice alerts over alphanumeric links or ATIS information. Comments received indicated that the low ranking of ATIS was due to the time between updates.

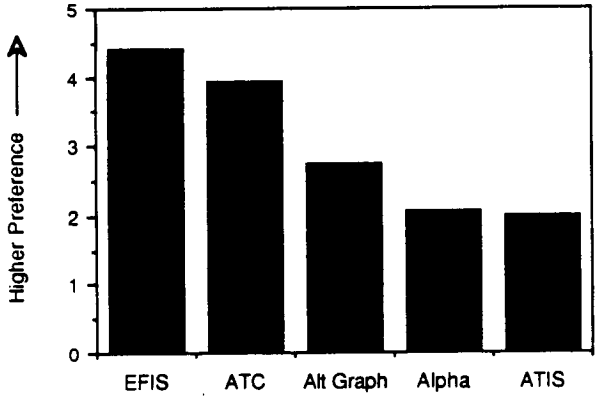


Fig. 9: Pilot rankings of possible relay/presentation modes of windshear information from the ground

A preliminary experiment has been conducted with a general aviation simulator to compare the efficiency of voice and graphical modes of presentation. Eight GA pilots with 210 to 1,700 total flight hours were tested. The scenario involved a microburst which appeared during an ILS approach, when the aircraft reached the outer marker. Avoidance of the microburst required a non-standard missed approach. The information was presented by voice, on a runway-fixed graphic display of microburst position, and on a graphic display showing both the microburst and aircraft positions. The data (Table 3) shows the effectiveness of the graphic displays. Avoidance improved significantly with the graphic displays, even though the same information was presented at the same time in each case.

Presentation Type	Avoidance Rate
Voice (JAWS format)	43%
Runway-Fixed Graphical Display: Microburst position only	62%
Runway-Fixed Graphical Display: Microburst + Aircraft position	94%

Table 3: Results of experiment with general aviation simulator and computer graphic display.

These results considered, the incident at Stapleton Airport serves as an illustration of the problems of voice communication. Crew and ATC workload in terminal phases of flight is high, leading to possible confusion and error. The simulation indicates that even under fairly light workload, the difficulty involved in fully interpreting the microburst threat from a voice warning can mean the difference between avoidance and penetration. Further evaluation of communication modes, including a variety of alphanumeric and graphic formats, will be performed with the part-task 757/767 simulator.

3.3.3 Message content and timing

The issue of what data is necessary and when it should be presented is important for either voice or digital transmission. In either case, a limited amount of information can be contained, and the timing must be determined to give the crew maximum awareness while minimizing the increase in workload. An initial viewpoint can be obtained from the pilot surveys. The responses indicate that location and intensity of microbursts are clearly the most important information items. Size, microburst movement, and intensity trends are of secondary importance, and shape data is generally felt to be inconsequential.

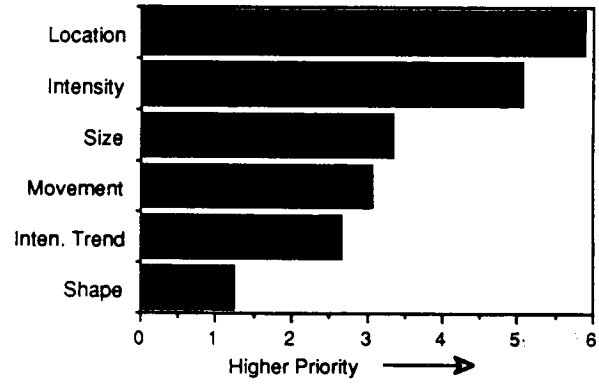


Fig. 10: Pilot ranking of microburst information by importance

The issues of what constitutes a hazardous microburst, who should be informed, and when are more difficult to resolve. The surveys are less clear in this case; the most common response was that aircraft should be alerted as soon as microbursts are detected anywhere around the airport vicinity. A few pilots defined a particular phase of flight, i.e. at the outer marker, when cleared for approach, or immediately upon entering the terminal area, as the best point for delivery of microburst alerts.

In response to a question about threshold shear levels, there was general agreement that a windshear advisory should be issued for approximately 10 knots of head-to-tail shear and a warning for 15 knots of shear. Also, it was almost unanimously expressed that decisions about the threat posed by windshear in a particular situation should be made entirely by the pilot, and the controller's role should be to maintain safe separation during

avoidance maneuvers. However, it remains to be determined what locations and intensities of microbursts actually constitute a threat in the view of the pilot. It is impractical to plan on distribution of all available windshear information in raw form to all aircraft in a congested terminal area. Some 'threshold hazard level' needs to be defined, based not only on the windshear intensity of the microburst, but including other factors such as the microburst and aircraft locations, aircraft altitude, and desired flight path.

3.4 Current Research

Research to resolve these issues is being conducted with the part-task 757/767 simulation shown in Fig. 11. The simulation uses an IRIS 2400T graphics computer, an autopilot control panel, and an EFIS control panel to duplicate the electronic instrumentation and flight dynamics of the aircraft. Data from TDWR experiments is used to generate simulated airborne weather radar returns and the windfield over the airport.

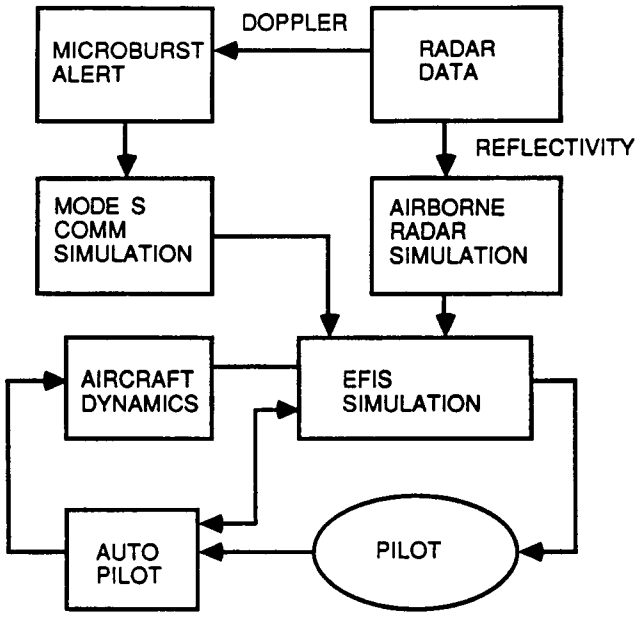


Figure 11: Part-Task 757/767 EFIS Simulation

The initial simulations are based on data provided by NCAR and the Lincoln Laboratory TDWR evaluations. The Stapleton incident is very well documented, and serves as a model for scenario construction. With a suitable sidetask, workload levels will be properly adjusted to get a reasonable range of pilot responses. Once the simulation is validated and a range of scenarios developed, issues of information format can be explored. Simulations of Mode-S transmissions with varying alphanumeric and graphic alert formats can be added, as well as voice communications, and the differences in pilot decision making and reaction time can be measured. If the results are commensurate with the results of the earlier general aviation simulations, more specific tests can be performed. These will center around more specific information issues, such as warning content, timing, and display formats.

4. Conclusions

Based on the above, the following points can be made:

- Technological advances in weather sensors and information transfer will allow development of sophisticated hazardous weather detection and alert systems. Design guidelines, centered around the end user and the available equipment, need to be applied to these systems.

- Microburst windshear is a weather hazard of particular concern and hence provides a good test case for development of user-oriented weather alert displays. Pilot surveys indicate that currently available detection and alert systems are not adequate, and a system for advance detection and alert is needed.

- The events which occurred during the TDWR operational evaluation on July 11, 1988 were analyzed in the context of information transfer issues. The observations included the following: Variability of pilot response to similar microburst alerts. The verbal relay of microburst alerts was found to induce delays. The inclusion of microburst alerts with other routine messages was thought to reduce the sense of urgency of the alerts. Finally, PIREPS have been found to be extremely important in validating the TDWR alerts to the user.

- A review of the current state of microburst detection technology and the analysis of the TDWR operational evaluation leads to the conclusion that the integration of ground-based doppler radar, LLWAS and PIREPS, and of a digital datalink such as Mode-S is the most viable near-term system for reliable advance warning of windshear.

- A simple flight simulator study has indicated that display of windshear information with a graphical display of aircraft and microburst position can result in significantly greater microburst awareness and greatly improve the probability of avoidance when compared with standard voice transmission.

- An opinion survey of air carrier pilots was conducted. Pilots feel that PIREPS and visual clues are the best currently available methods for microburst detection, while LLWAS and airborne weather radar are less effective. Also, pilots were receptive to the idea of displaying windshear information on an EFIS display, preferring the EFIS to ATC voice communications. Alphanumeric information and ATIS were rated poorly for transmission of windshear alerts.

- When asked about microburst alert information content, pilot specified that microburst location and intensity were the most important items, followed by size, movement, and intensity trend information.

- Research is currently in progress to further explore the issues involved. A part-task Boeing 757/767 simulation has been developed to address issues of warning content, timing, and a selection of alphanumeric and graphical display formats.

Acknowledgments

This work was supported by the MIT Lincoln Laboratory under contract BARR-10-119, and the Federal Aviation Administration and the National Aeronautics and Space Administration under grants NGL-22-009-640 and NAG-1-690. Data from the 1988 TDWR operational evaluation was provided by Dr. Wayne Sand and Cleon Biter of the NCAR Research Applications Program. The authors would like to thank the respondents, Rick Brown, and United Airlines for assistance with the pilot opinion surveys.

References

1. National Research Council, *Low Altitude Wind Shear and Its Hazard to Aviation*, National Academy Press, 1983.

2. Targ, R., and Bowles, R.L., "Investigation of Airborne Lidar for Avoidance of Windshear Hazards," AIAA Paper 88-4658, Sept. 1988.

3. Bracalente, E.M., Britt, C.L., and Jones, W.R., "Airborne Doppler Radar Detection of Low Altitude Windshear," AIAA Paper 88-4657, Sept. 1988.

4. Adamson, H.P., "Airborne Passive Infrared System for the Advance Warning of Low-Level Windshear and Clear Air Turbulence: 1988 In-Service and Theoretical Work," AIAA Paper 88-4659, Sept 1988.

5 Wolfson, M.M., "Characteristics of Microbursts in the Continental United States," *The Lincoln Laboratory Journal*, Vol. 1, No. 1, Spring 1988, pp. 49-74.

6. Federal Aviation Administration, *Windshear Training Aid*, 1987.

7. National Center for Atmospheric Research, "Terminal Doppler Weather Radar (TDWR): A Briefing Paper," July 1, 1988.

8. Private Communication with McEachen, A. of Aeronautical Radio, Inc., August 1988.

9. Rucker, R.A., Flathers, G.W., "The Future for Aeromobile Digital Communications," *IEEE/AIAA 8th Digital Avionics Systems Conference*, October, 1988.

Time and Truth in Plans*

Lyman R. Hazelton, Jr.
Flight Transportation Laboratory
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology

1 Introduction

Planning denotes the formulation of a detailed scheme, program, or method worked out beforehand for the accomplishment of a goal. It involves the analysis of the desired goal and its division into sub-goals which are subsequently treated in the same way until a set of primitive objectives is obtained. A rational plan is prepared by a reasoner for execution by one or more actors or agents who perform actions to achieve the objectives. A reasoner is a cognitive system (human or machine) capable of some level of logical deliberation.

The formulation of complex plans is an arduous process. Computers have been employed to assist in the creation of plans almost from their inception. One of the first uses of computers, shortly after World War II, was to solve large linear optimization problems for military planners. However, the actual creation of a plan by a computer did not occur until Sussman's work [1] in 1975. The reason for this delay is that planning is a cognitive process not directly involving computation; that is, planning requires symbol manipulation. Also, the composition of a plan is highly domain specific, so it is difficult to make a general planning program [2]. Domain dependent plan generation is *rule based*, and the rules vary from domain to domain. The creation of plans by a computer had to await the development of a mature symbol manipulation language, such as *Lisp*.

2 The Single Actor Assumption

Sussman, and most of those who followed him, made a very powerful simplifying assumption regarding the domain of a planner. This assumption is that all actions in the domain are taken by a single actor, the plan executor. Furthermore, that actor is

*Research supported by NASA contract number NGL 22-009-640.

restricted to taking only one action at a time. This supposition constrains the plan to consist of a *serial sequence* of primitive objectives. Because the sequence is non-overlapping, there is no requirement for the planner to be aware of time in the general sense. Time in this restricted domain is only measured by the conclusion of a task; the clock time necessary to complete the task is unimportant. The assumption also guarantees that the environment is not hostile. Because there is only one actor, and it carries out the plan, there is no mode by which the plan may fail. This obviates *execution monitoring* and plan *repair*.

The central reason for the employment of the single actor assumption is the lack of a unified representation of the knowledge required by the more complex environment. The standard representations employed by classical planners are sufficient to represent domain knowledge alone. Information concerning causal or temporal relationships is not domain knowledge in the normal sense. Rather, this kind of information is an example of more abstract knowledge regarding relationships between facts known about the domain. For example, let us analyze the statement, "After snow plowing is completed, the runway will be returned to service." This statement is actually *three* statements: (1) Snow plowing of the runway will be completed, (2) the runway will be returned to service, and (3) the period of validity of statement (2) follows that of statement (1) temporally. Notice that statement (3) is a statement about statements (1) and (2), without regard to their domain meaning or content. Making such meta-statements requires a mechanism within the representation language which allows reference to *statements* that have previously been made.

The domain of classical planners also provides the reasoner with accurate information about the initial state of the domain and a precise goal in terms of the required final state.

3 The Real World

Although such restricted environments do exist in reality, they are very rare and usually contrived. In the "real world", there are multiple actors who can accomplish tasks simultaneously. Some of the actors may be random or chaotic or even hostile to the planner — therefore a plan or a part of a plan may fail. There may be clock time constraints on the accomplishment of a task, as well as relative temporal constraints *between tasks.* Additionally, "real world" plans may depend on causality, a notion completely missing in classical computer planners. A planner in the real world must not only *create* a plan. It must *schedule*[1] the events of the plan, *monitor* the execution of the plan and detect failures, and *repair* the plan and schedule in light of any failures that occur.

In "real world" environments, domain information may be incomplete or erroneous. The planner is required to discover spurious data and to be able to make reasonable assumptions about missing or incorrect knowledge. Finally, "real world" planners often must operate in a process management domain, in which there is *no* final goal, but rather a continuous requirement to control and keep a system "functioning". A planner with such a task must analyze the current and predicted future states of the system and even *create* goals.

4 Forward Chaining Inference

Computer reasoners are based on *rules.* For the purposes of this discussion, a rule is a generalized statement of deduction. Rules are composed of two parts, an antecedent consisting of one or more conditions, and a consequent consisting of one or more declarative statements regarding changes in the reasoner's knowledge. The

[1]Scheduling is the assignment of specific times to the objectives determined by a planning process.

consequent of a rule is valid if and only if all of the conditions in its antecedent are valid.

When a new statement about the reasoner's knowledge is asserted, the reasoner examines the antecedents of its rules to see if any of them are applicable. Those that are deemed appropriate are queued (together with contextual information, in a structure called a situation) for more detailed examination on the reasoner's agenda. When the reasoner completes its analysis of a rule it gets the next rule from the agenda, until the agenda is empty. The reasoner then has an up to date picture of the domain until some new observation is perceived.

5 Truth Maintenance

Knowledge about the world can be classified as *observed* fact and *inferred* fact.

Observed fact is information provided to the reasoner from some direct observation of the domain.

Inferred fact is knowledge produced by the reasoner via the application of its rules.

As its observations of the world change, a reasoner must change its inferred beliefs in order to preserve a correct representation of reality. To accomplish this, the reasoner must retain information regarding the evidence supporting each fact that it believes. In practical applications, this information is kept in a tree structure. The nodes of the tree represent known facts, while each arc represents an inference relation or *dependency* (see Figure 1).

The base nodes in the tree represent observed facts. All other facts in the tree are inferred facts. When an observation changes, all the inferred facts reachable by

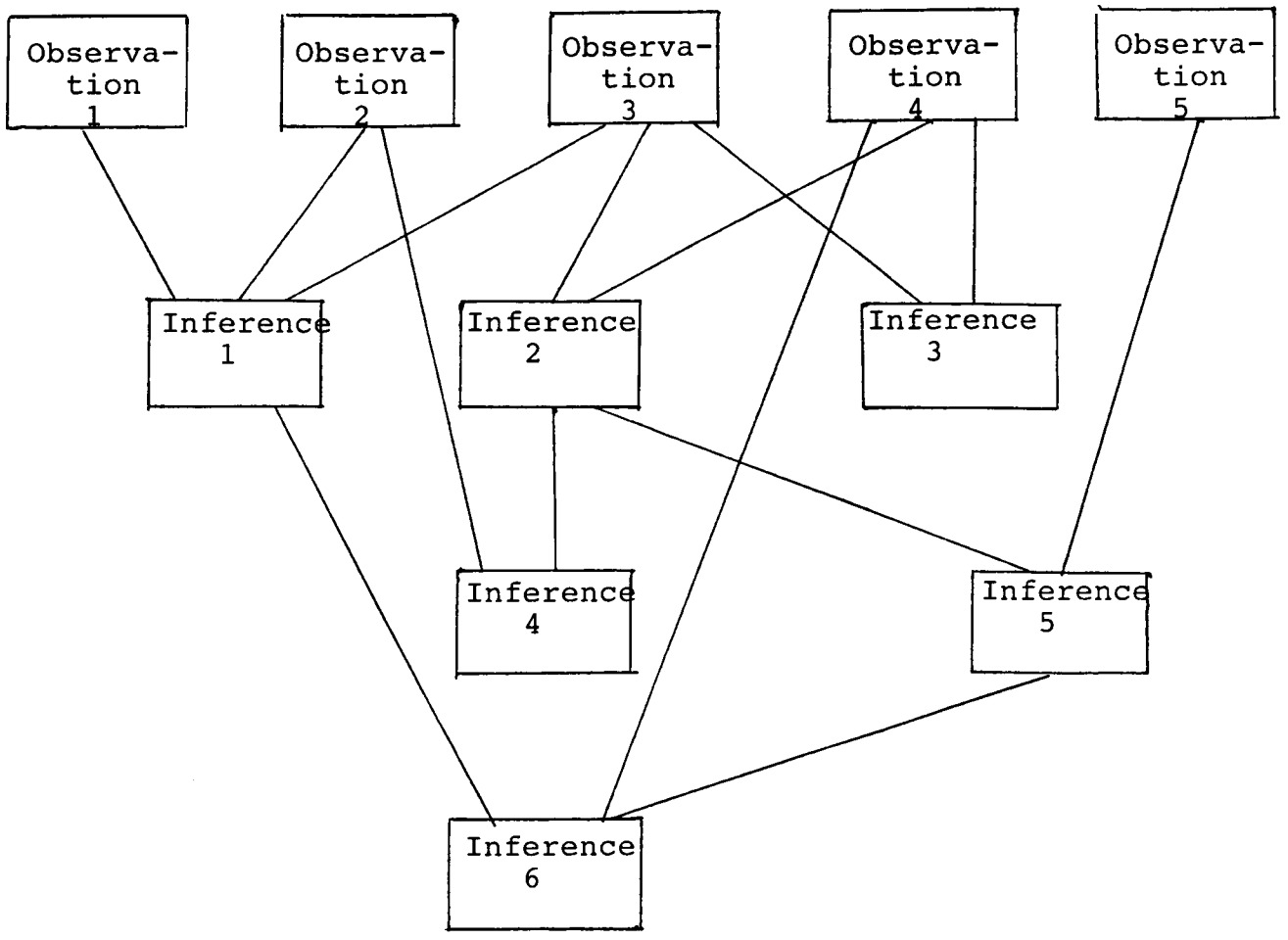


Figure 1: Evidentiary Tree

going *down* dependency arcs from the modified observation node must be investigated for correctness. Those found to be incorrect under the new observation must either be corrected or denied (i.e., deleted from the tree). All inference nodes which are dependent on a changed or denied node must likewise be inspected.

The process of maintaining a correct set of beliefs through the use of evidence is called *Truth Maintenance* ([3] and [4]). It should be noted that the process of truth maintenance is a form of *constraint propagation.* It is said that the truth of a node is constrained by the truth of its supporting nodes. The validities of all inferred facts are constrained by the validity of the facts that were used as a basis for their deduction. The inspection and update process propagates down the tree. The system of programs that implements this constraint propagation process is called a *Truth Maintenance System* or *TMS*.

6 Assumptions and Truth Maintenance

When a reasoner is faced with incomplete information, it must resort to making tentative assumptions about the questionable domain. The reasoner then proceeds to infer the consequences of those assumptions on its representation of reality. This is another classification of knowledge:

> *Assumed fact* is knowledge "guessed" in order to proceed with a chain of reasoning when the reasoner detects that it is missing some information.

A *contradiction* occurs whenever a rule deduces a fact that conflicts with another fact already known to the system. Should a contradiction be detected, the reasoner is forced to *change* one or more assumptions. Changing an assumption requires truth maintenance similar but not identical to that required when changing an observation.

Doyle [3] implicitly defines all non-deduced knowledge in his truth maintenance system as assumption. There is a disadvantage to this view. It unnecessarily increases the number of candidates when a contradiction is found. I prefer to view information that the reasoner *observes* as different than that which is *assumed* because of some lack of knowledge.

The process of changing an assumption when a contradiction is detected requires two steps. The first is to *find* an appropriate candidate to change. This step, first explored by Stallman and Sussman [5], is called "dependency directed backtracking". The second step is to remove the offender and replace it with a different assumption, if possible.

An important question needs to be addressed when reasoning with assumptions: *When* is it appropriate to make an assumption? Clearly, just because we *can* make an assumption does *not* mean that we *should* make it. Making unnecessary assumptions is expensive and can be confusing. We need to have a criterion which determines the *need* to make an assumption. I claim that the only time that we need to make an assumption is when we want to know some missing information. Thus, when we are in the middle of deducing something new about our world and we find that we are missing some fact required in the chain of reasoning, we try to make an assumption about what is unknown that is congruous to what we do know. Operationally, this means that we need a *special* set of rules for making assumptions.

These assumptive rules have the same general form as normal rules. However, they are different in three significant ways.

1. They are only invoked during the evaluation of the antecedent of a normal rule. This insures that assumptions are only made when they are required.

2. If all of the antecedents of an assumptive rule are satisfied, then the consequent asserts the assumption through the TMS, as usual. However, it is *not* necessary to search for other rules which will use the assumption and enqueue situations on the agenda. This is because if there were another rule requiring the assumption, the assumption would already have been made.

3. Assumptive rules can be *disabled* or *enabled* by the system. There are circumstances under which the reasoner can detect that a particular assumption cannot logically be made, and the associated assumptive rule(s) should not even be considered. If the circumstances change, the restriction may be lifted.

To understand why assumptive rules should be disabled, consider what must take place if an assumption results in a contradiction. First, note that contradictions can occur for two reasons:

1. Suppose that facts A and B are used by rule R_1 to deduce fact C. Now suppose that the resaoner observes that **not** C is true, and that A and B are true. In this situation, either the user is mistaken about one of the facts, or the rule R_1 is incorrect. There is no way for the reasoner to determine which of these premises is true, so it should stop and explain the situation and ask for a correction.

2. Suppose, again, that facts A and B are used by rule R_1 to deduce fact C. Now suppose that the reasoner observes that A is true, and that the system has no knowledge about B. The fact that A is true will cause the reasoner to attempt to run rule R_1. When the antecedent requiring B is evaluated, the reasoner will come up empty handed and then endeavor to make an assumption about B. If there is an appropriate assumptive rule AR_1 which succeeds, assumption B' is

asserted. The C' resulting from $\mathcal{R}_1$ will *also* be an assumption. Now, if **not** C' is true, a contradiction will occur.

Contradictions involving assumptions need to be looked at in greater detail. In the example above, clearly the reasoner should deny both C' and B'. In addition, it should mark the assumptive rule $\mathcal{AR}_1$ as unusable (i.e. disabled) as long as **not** C' is known to be true. It can then consider other assumptive rules that may result in an alternative hypothesis for B. If there are no more assumptive rules for B, then processing of $\mathcal{R}_1$ must be terminated.

7 Time and Truth Maintenance

The truth maintenance systems of Doyle and DeKleer do not concern themselves with an environment that changes with time. In these systems, either a statement is supported by current data, or it is not.

When new data are obtained, truth maintenance systems change the implications that were previously derived based on old knowledge. Consider the following rules for a resource allocator:

1. You may only allocate a resource if it is not already allocated (in use).

2. If you do not know that a resource has been allocated, you may assume that it is available (not allocated).

In a system with truth maintenance, if the allocator wants to assign a resource, it looks for one which is uncommitted. Suppose that it only finds resources that are already committed or about which it is ignorant. It then *assumes* that one of the uncommitted resources is available. Based on the evidence of that supposition, the resource can be allocated. When the resource is allocated, the assumption that it is available is no

longer true, and the TMS removes it. But the assumption of availability was the evidence supporting the action of allocating the resource in the first place, so it must remove the fact that it has allocated the resource, too. This is a variant of the famous statement, "This statement is not true." The system wants to state that the resource is available and unavailable at the same time. While the TMS may be able to detect circularities of this kind, it cannot properly resolve them without modification.

The reason that the TMS gets into trouble in the resource allocation problem is that it has no temporal knowledge. All of its statements must be true at all times. To correct this difficulty, it must be able to reason not only *why* a statement is true, but *when* it is true, as well.

Interest in temporal reasoning has increased recently. Notable work has been done by James Allen ([6] and [7]), Tom Dean ([8] and [9]), and Yoav Shoham ([10] and [11]). Prior to their work, the most that one could do was to "time tag" each fact in the system with the clock time at which it was asserted.

The real break through in Allen's work is the idea that most temporal references about facts are *relative* to temporal references of other facts. Allen further postulates that temporal references are not points in time, but intervals. That is, there is a period of time, which I like to call the "validity interval" of the fact, during which the fact is true.

Very few, if any, facts are actually tagged with a clock time in Allen's system. Instead, most validity intervals are determined by constraints derived from relationships to other facts. For example, two facts may be connected by an "after" relation: "The runway may be returned to service *after* the plowing is completed." The significance of this connection is that if the period of validity of the plowing operation is extended (i.e., the plowing takes longer than was expected), the runway must be returned to

service later. The period of validity of the plowing operation constrains the validity period of the return to service. The result is another constraint propagation network, but a rather complex one. While it has been shown [12] that the maintenance of the full temporal network described by Allen is at least NP-hard, the approach has interesting implications to truth maintenance systems.

In the truth maintenance constraint propagation process described above, the arcs in the network identify a single relationship: implication. According to Allen, there are thirteen different ways that one temporal interval can be related to another. Further, there is a transitive algebra that allows one to compute the relationship between two intervals that have known relationships to a common interval. This means that it is not necessary to keep *all* of the arcs that could connect the nodes in the net, but only a sufficient number of them to allow the others to be computed as necessary. The computation of a transitive relation is claimed to be fast, so it seems reasonable to keep only a minimum network.

Every fact known to the system must have an associated temporal validity interval. Facts which are true at all times have a special temporal interval "ALWAYS". While we could also support facts which are never true with another special interval, it is more efficient to simply exclude them all together. Note that a statement of the form "X is never true" is a true statement over the interval ALWAYS.

In a planning domain, it is not enough to simply change the rules to incorporate temporal intervals and use Allen's transitive operators. To see this, consider the resource allocator previously discussed. In a temporal environment, the rules must be modified to include validity interval information:

1. You may only allocate a resource during a time interval $\mathcal{T}$ if it is not already allocated at any time during $\mathcal{T}$.

C-2

2. If you do not know that a resource has been allocated during a time interval $\mathcal{T}$, you may assume that it is available for allocation during $\mathcal{T}$.

All that this addition does is to restrict the interval of time from all time to the interval $\mathcal{T}$. While it is a necessary addition to the rules for operation in a temporal environment, the reader will see by stepping through the previous exercise that it is not sufficient to solve the problem.

The solution to the problem can be realized by observing the logic employed by a human scheduler. The human is aware of *two* times while constructing the schedule. One of the times is the period during which the resource is to be allocated, called $\mathcal{T}$ above. The other time is the actual time that the plan is being constructed, the planner's *now*. The planner is not only aware of the time sequence of the proposed schedule, but of the sequence of events during plan construction. In a temporal environment, contradiction can only happen between facts which have overlapping temporal validity intervals. In the same way that new information can change the truth of implied facts in the TMS, new information can change the temporal validity intervals of facts that were already known in the temporal system. This process goes beyond simply computing the relationship between two facts in the network.

As an example, let us re-examine the resource allocation problem once again. Suppose the allocator wants to assign a resource during a time interval $\mathcal{T}$. It begins by looking for an uncommitted resource. Suppose, as previously, it finds only resources which are committed during $\mathcal{T}$ or about which it is ignorant concerning allocation status during $\mathcal{T}$. It can then *assume*, at time τ_1, that one of these unknown resources is available during $\mathcal{T}$. Note that τ_1 is the time that the availability assumption is made, and has *nothing* to do with the proposed schedule time $\mathcal{T}$. The temporal validity interval of the assumption is from τ_1 to forever. On the basis of the assumption, the

resource can be allocated during $\mathcal{T}$ at time τ_2, which is later than τ_1. This allocation produces a fact that the resource is not available for the period $\mathcal{T}$, with a temporal validity interval from τ_2 to forever. The effect of this new fact is not, however, to completely remove the assumption that the resource was available from τ_1 to forever. Rather, it removes the validity of that assumption from τ_2 to forever. This leaves the assumption that the resource is available valid from τ_1 to τ_2. Thus the evidence for the allocation step remains valid.

This mechanism is still a form of constraint propagation, since the new information about the plan constrains the temporal validity interval of old plan information. While it is possible that this change may effect the validity intervals of yet other facts, in practice it does so only rarely.

8 Conclusions

In this report, I have described some of the logical components of a rule based planning and scheduling system. I have pointed out a deficiency in the conventional truth maintenance approach to this class of problems and suggested a new mechanism which overcomes the problem.

This extension of the idea of justification truth maintenance may seem at first to be a small philosophical step. However, it embodies a process of basic human reasoning which is so common and automatic as to escape conscious detection without careful introspection. It is vital to any successful implementation of a rule based planning reasoner.

References

[1] Sussman, G. J., *A Computer Model of Skill Acquisition*, Elsevier, 1975.

[2] Chapman, D., *Planning for Conjunctive Goals*, M.I.T. Artificial Intelligence Laboratory Technical Report 802, 1985.

[3] Doyle, J., *A Truth Maintenance System*, Artificial Intelligence, Volume 12, 231-272, 1979.

[4] deKleer, J., *An Assumption-Based Truth Maintenance System*, Artificial Intelligence, Volume 28, 127-162, 1986.

[5] Stallman, R. M., and Sussman, G. J., *Forward Reasoning and Dependency Directed Backtracking in a System for Computer Aided Circuit Analysis*, Artificial Intelligence, Volume 9, 1977.

[6] Allen, J. F., *Maintaining Knowledge About Temporal Intervals*, Communications of the ACM, **26**(11), 832-843, 1983.

[7] Allen, J. F., and Kooman, J. A., *Planning Using a Temporal World Model*, Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI-1983, 1983.

[8] Dean, T., *Intractability and Time-Dependent Planning*, Proceedings of the 1986 Workshop on Reasoning about Actions and Plans, 245-265, Morgan Kaufmann, 1987.

[9] Dean, T., and Kanazawa, K., *Probabilistic Causal Reasoning*, The Fourth Workshop on Uncertainty in Artificial Intelligence, 73-80, 1988.

[10] Shoham, Y., *Reasoning About Change*, M.I.T. Press, Boston, MA, 1987.

[11] Shoham, Y., *Time for Action*, Proceedings of the International Joint Conference on Artificial Intelligence, 954-959 & 1173, 1989.

[12] Kautz, H., and Vilain, M., *Constraint Propagation Algorithms for Temporal Reasoning*, Proceedings of the Fifth National Conference on Artificial Intelligence, AAAI-1986, 1986.

OHIO UNIVERSITY

PRECEDING PAGE BLANK NOT FILMED

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT OHIO UNIVERSITY 1988-1989

Richard H. McFarland
Department of Electrical and Computer Engineering
Ohio University
Athens, Ohio

INTRODUCTORY REMARKS

This year has been an especially exciting one at Ohio University with engineering research into three very contemporary issues. Not only does the technical content stir excitement in the lives of young engineers-in-training but also it provides for our more senior staff an opportunity for a critical insight into topics that are of current interest to our society. This judgment is made based on the kind and amount of response that has been received at technical meetings and the interest in government concerning the new technical findings.

The first major topic is the use of the NAVSTAR Global Positioning System (GPS) for purposes other than the original military application. Because Ohio University has been pursuing research for many years in the field of air navigation by using Loran-C it was logical that it would be integrated into the thinking of those here at Ohio who were specializing in GPS navigation. Combining these two areas into one involving interoperability was quite logical and natural. This is especially true when one considers that there are certain deficiencies known to exist with the GPS which need early solution for this expensive system to have maximum societal benefit. These deficiencies, monitoring for integrity and incomplete global coverage challenged technical staff and students to come up with solutions without adding substantial costs. A good solution will ultimately allow GPS and Loran to be considered and designated as a hybrid sole means of air navigation.

Work was performed by Frank van Graas to show an effective complementary structure. This work was progressively reported over the last two years and serves as the principal element in his Doctoral Dissertation completed in November 1988.[1] The originality of the work was documented in his article in the Summer 1988 Issue of the Journal of The Institute of Navigation.[2] The work is

significant enough that national interest in GPS/Loran interoperability has developed at least in part because of the positive results of van Graas' work.

The second major area of work which has carried its own excitement with it, especially with pilots who fly regularly, is that of uplinking of near real-time weather data to the cockpit. Adverse weather can be life threatening. The pilot is the one person who is in fact and by regulation given the responsibility of the safety of the flight. Yet, he is the person who sometimes has the least amount of meaningful information. Considerable weather data are presently being generated by the Federal Aviation Administration (FAA) and the National Weather Service. Newer and better radars, for example, are providing more complete and detailed weather products which are essentially unavailable to the pilot in flight when he needs these most. Some of the best data comes in graphical form and there is no provision to get this to the cockpit where the decision is made.

Perhaps it is better said that without the product there would be more excitement in the back of the aircraft and in the cockpit, too. Excitement admittedly can come in various forms, but the weather data uplink research provided some true technical excitement when Craig Parker presented his results[3] from his Masters Thesis project. He demonstrated an impressive data compression for weather data which makes it quite realistic to place this data stream on available, aircraft common very-high frequency (VHF) communication channels.

The potential benefits available from Parker's work are many. More extensive, complete data, some in graphical form being available to the pilot, can only translate into higher levels of aeronautical safety. The well-recognized 40% figure of aircraft accidents related to weather can be expected to decrease with implementation of weather data uplinks. The issue of how to handle the mass of available data is addressed by Parker. More information bits per unit time per channel is what is needed and the thesis covers several possibilities. One of these, quadrature phase shift keying was demonstrated as part of the thesis program with an uplink to an aircraft in flight using a Piper Saratoga plus some live link work for the conference room during the thesis defense.

This FAA/NASA supported program has made it possible for these important projects to exist. High quality technical results have been derived from this work, and quite importantly for Ohio University, the students involved, and society as a whole, have derived benefits which should prove to be extraordinary.

High Speed Civil Transport (HSCT) technology is being addressed now and the work product will be discussed and reported in the near future. In this topic, as

with the other more mature topics of investigation, both undergraduate and graduate students have been involved along with four Electrical and Computer Engineering faculty members. A total of five students have participated over this past year. Indications from prospective employers are that this program is an important item helping establish credibility for new engineers. Ohio University strongly believes this program is a very important part of its educational program in engineering.

Evidence of the productivity of the program can be obtained by referring to the annotated bibliography which follows.

ANNOTATED REFERENCES OF 1988 PUBLICATIONS

1. van Graas, F.: Hybrid GPS/Loran-C: A Next-Generation of Sole Means Air Navigation. Ohio University, Department of Electrical and Computer Engineering, Dissertation, November 1988.

This paper describes a new technique that hybridizes the NAVSTAR Global Positioning System (GPS) and the Long Range Navigation system, Loran-C, based on a generic pseudorange processing technique. The concept, theoretical analysis, and justification of a hybrid GPS/Loran-C system are presented, along with a scheme for meeting sole means of navigation requirements. Following the design and modeling phase, a prototype GPS/Loran-C receiver was developed and implemented. The hybrid GPS/Loran-C receiver concept was proved through an actual flight-test, which was referenced to a Differential GPS truth trajectory. The hybrid system has the potential to meet all requirements for a next generation of sole means of air navigation for the conterminous United States.

2. van Graas, F.: Sole Means Navigation Through Hybrid Loran-C and GPS. *Navigation: Journal of The Institute of Navigation*, Vol. 35, No. 2, Summer 1988.

A minimum of four GPS range measurements or two Loran-C time differences is normally required for a position solution for en route navigation, area navigation, and nonprecision approaches. This paper describes a new technique that hybridizes GPS and Loran-C used in the pseudorange mode to process efficiently all available navigation information. Emphasis is placed on combined GPS and Loran-C timing, both for the ground/space facilities and the user equipment. The hybrid system has the potential to solve the GPS and Loran-C integrity and availability problems; more measurements are available than are required for the navigation solution.

3. McFarland, R.H. and Parker, C.B.: Weather Data Dissemination to Aircraft. *Proceedings of the AIAA 27th Aerospace Sciences Meeting*, AIAA Paper 89-0809, Reno, Nevada, January 9-12, 1989.

Documentation exists that shows weather to be responsible for approximately 40 percent of all general aviation accidents with fatalities. Weather data products on the ground are becoming more sophisticated and greater in number. Although many of these data are critical to aircraft safety, they currently must be transmitted verbally to the aircraft. This process is labor intensive and provides a low rate of information transfer. Consequently, the pilot is often forced to make life-critical decisions based on incomplete and outdated information.

Automated transmission of weather data from the ground to the aircraft can provide the aircrew with accurate data in near real time. The current National Airspace System Plan calls for such an uplink capability to be provided by the Mode S beacon system data link. Although this system has a very advanced data link capability, it will not be capable of providing adequate weather data to all airspace users in its planned configuration. This paper delineates some of the important weather data uplink system requirements, and describes a system which is capable of meeting these requirements. The proposed system utilizes a run-length coding technique for image data compression and a hybrid phase and amplitude modulation technique for the simultaneous transmission of both voice and weather data on existing aeronautical Very High Frequency (VHF) voice communication channels.

RIDGE REGRESSION PROCESSING

Mark R. Kuhl
Ohio University
Athens, Ohio

SUMMARY

Current navigation requirements depend on a geometric dilution of precision (GDOP) criterion. As long as the GDOP stays below a specified value, navigation requirements are met. The GDOP will exceed the specified value when the measurement geometry becomes too collinear. A new signal processing technique, called Ridge Regression Processing, can reduce the effects of nearly collinear measurement geometry; thereby reducing the inflation of the measurement errors. It is shown that the Ridge signal processor gives a consistently better mean squared error (MSE) in position than the Ordinary Least Mean Squares (OLS) estimator. The applicability of this technique is currently being investigated to improve the following areas: receiver autonomous integrity monitoring (RAIM), coverage requirements, availability requirements, and precision approaches.

BACKGROUND

Ridge Regression was developed by statisticians Hoerl and Kennard in 1970 (ref. 1). Throughout the 70's Ridge Regression was a controversial topic in the statistics world as shown by Efron (ref. 2). This was because Ridge Regression is a biased estimation technique - it goes against the traditional unbiased Ordinary Least Mean Squares (OLS) technique which has been used since Gauss. Based on the statistician's Ridge Regression techniques, a Ridge signal processor was developed in 1988 by Kelly (ref. 3) for navigation applications.

A navigation system gives a number of measurements that one tries to find system states (i.e. position, velocity, etc.) from by using a linear measurement model. The linear measurement model is given by the following equation:

$$\underline{Y} = H\underline{\beta} + \underline{e}$$

where:

$\underline{Y}$ is the measurement vector
H is the data matrix
$\underline{\beta}$ is the system state vector
$\underline{e}$ is the measurement noise vector

If no noise were present in the measurements, $\underline{\beta}$ could be solved for directly by taking the inverse of H and multiplying it by $\underline{Y}$. But since measurement noise exists, $\underline{\beta}$ must be estimated.

The OLS solution is given by the following equation:

$$\underline{\beta}_{OLS} = [H^T H]^{-1} H^T \underline{Y}$$

The GDOP, which is a factor that relates range errors and position errors, may be calculated by taking the TRACE of the first term of this equation, $(H^T H)^{-1}$. Looking more closely at this first term, one can see that when the values of the diagonal terms of the $H^T H$ matrix get smaller, GDOP increases. The second term of the equation, $H^T \underline{Y}$, is known as the smoothing function.

The Ridge Regression solution is given below:

$$\underline{\beta}_R = [H^T H + \kappa I]^{-1} H^T \underline{Y}$$

Again, the inflation of the position errors is calculated by taking the TRACE of the first term of the equation. But notice that it now includes a variable, κ, which is added to the diagonal elements of the $H^T H$ matrix. This variable, κ, limits the minimum values that the diagonal elements of this matrix can obtain, thereby limiting the inflation of the position errors to some maximum value. By limiting the inflation of the position errors, a better position estimate may be obtained.

RIDGE REGRESSION PROCESSING CONCEPTS

For navigation, the most important factor is the mean squared error (MSE) in position. The MSE grows as the GDOP grows. In order to understand GDOP, a look must be taken at the so-called "geometry problem". For example, if two range

measurements have a crossing angle of 90 degrees, the geometry is said to be "good"; but if the crossing angle between two range measurements is small, the geometry is said to be "bad". This geometry problem is embodied in the GDOP factor which, when using the OLS technique, is as low as 1.4 for good geometry (crossing angle is 90 degrees) and as high as 86 for bad geometry (crossing angle is 1 degree).

When bad geometry exists, the OLS technique inflates the variance of the errors around the true solution in the form of an error ellipsoid. The Ridge Regression technique adds a small bias to the true solution which in turn reduces the variance greatly, thereby giving a smaller MSE. Recall that the MSE is given by the bias term squared plus the variance, see figure 1.

A simulation is given for the Distance Measurement Equipment (DME) system with an aircraft, traveling at constant velocity, making range measurements to two DME stations. This simulation compares the performance of the OLS batch estimator to the Ridge Batch estimator. The simulation set-up is shown in figure 2. From figure 2, gamma is defined to be the crossing angle between two measurements. Notice that the error ellipsoid around the aircraft is a strong function of gamma (along with the measurement noise).

The simulation results are given in figure 3. Figure 3 compares the OLS estimator to the Ridge estimator for a crossing angle, gamma, of 1 degree. This figure shows graphically how the OLS estimator is centered on the true solution but allows the variance of the errors to grow greatly for small crossing angles. The Ridge estimator reduces this error variance, while introducing a small bias, giving an area of errors which is off from the true solution but having an overall smaller MSE (by a factor of 100!). The reduction of the MSE is achieved by choosing an appropriate bias term, κ. (Note that when κ is equal to 0, the Ridge estimator becomes the OLS estimator).

FUTURE RESEARCH

The effectiveness of receiver autonomous integrity monitoring (RAIM) for the Global Positioning (GPS) should be significantly improved by using the Ridge Regression Processing technique (ref. 4). Navigation system integrity is defined as the detection of a signal failure and a warning to the pilot that the system is not operating within required performance limits. (The FAA requires that the pilot be notified within 10 seconds during a nonprecision approach of a signal failure for a system to have integrity.) Right now, because there are times when bad geometry

exists among the GPS satellite sub-sets, integrity monitoring cannot be performed about 15% of the time. With Ridge Regression Processing used in place of the current OLS solution, GPS monitoring performance should improve.

Another avenue that needs to be pursued is the implementation of a recursive Ridge processor for precision approach applications. For precision approaches, very accurate positioning and fast output of the guidance information to the pilot is needed. The current Ridge batch processor used in the above DME simulation may give better position estimates when used in cooperation with the Kalman filter. But being a batch process, it may introduce too much lag in the guidance information. Therefore, the implementation of a recursive Ridge processor is desirable.

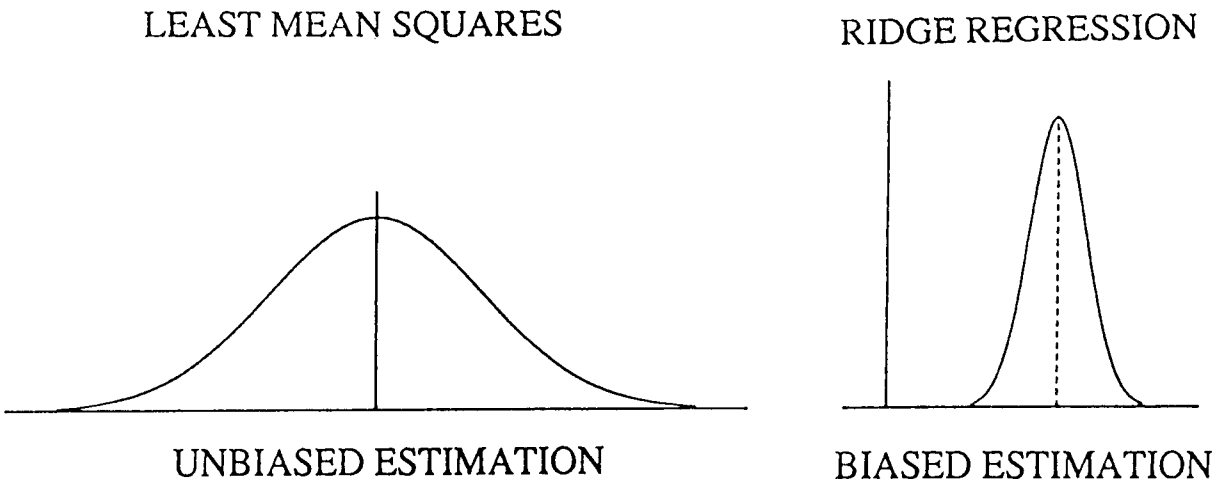
REFERENCES

1. Hoerl, A. E., and Kennard, R. W.: Ridge Regression: Applications to Nonorthogonal Problems. Technometrics, Vol. 12, No. 7, February 1970.

2. Efron, B.: Biased Versus Unbiased Estimation. Advances in Math, 16, 1975.

3. Kelly, R. J.: Reducing Geometric Dilution of Precision Using Ridge Regression Signal Processing. Proceedings of the IEEE PLANS'88, Orlando, Florida, November 29 - December 2, 1988.

4. Kelly, R. J., Van Graas, F., and Kuhl, M. R.: Improved Effectiveness of GPS RAIM Through Ridge Regression Signal Processing. Proceedings of The Institute of Navigation GPS 89 2nd International Technical Meeting, Colorado Springs, Colorado, September 27-29, 1989.



$$MSE = \quad BIAS^2 + VARIANCE$$

Figure 1. Comparison of unbiased and biased solutions and definition of Mean Squared Error (MSE).

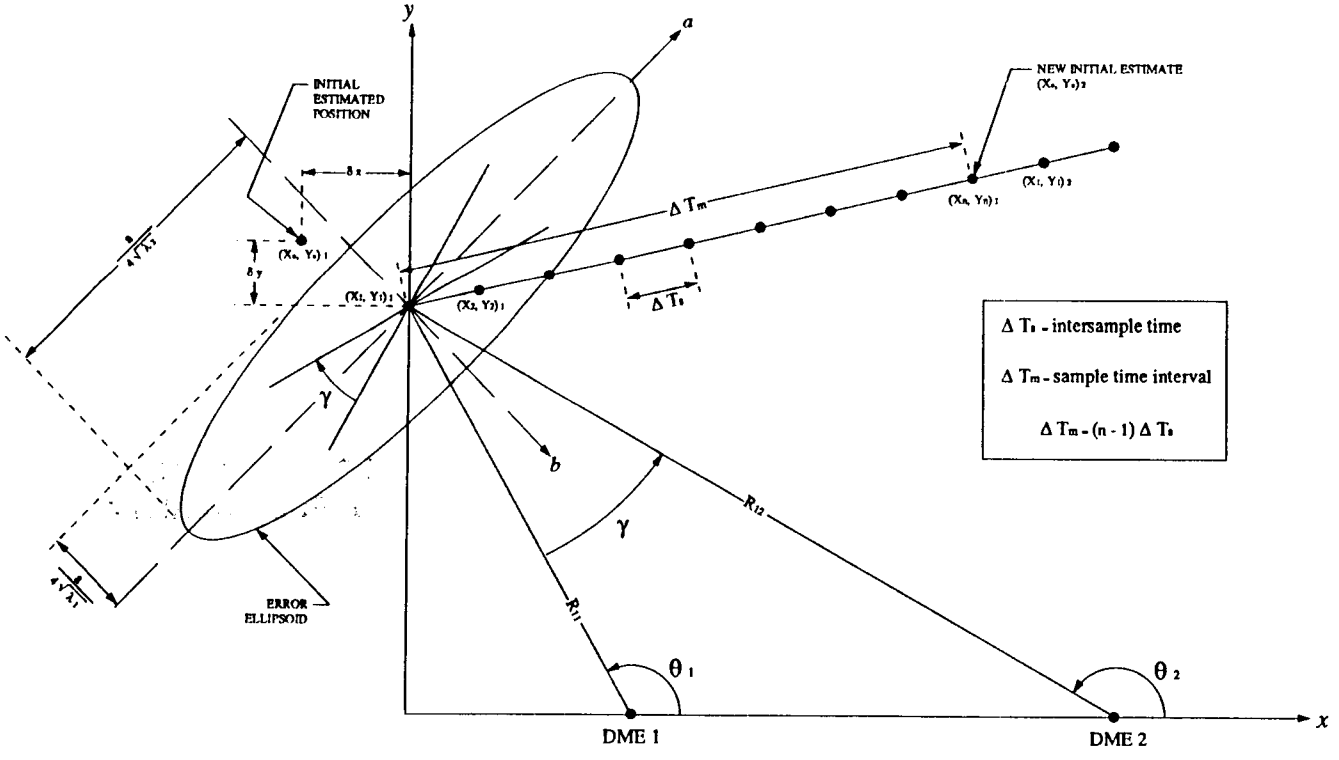


Figure 2. DME simulation geometry with an aircraft making range measurements to two DME stations.

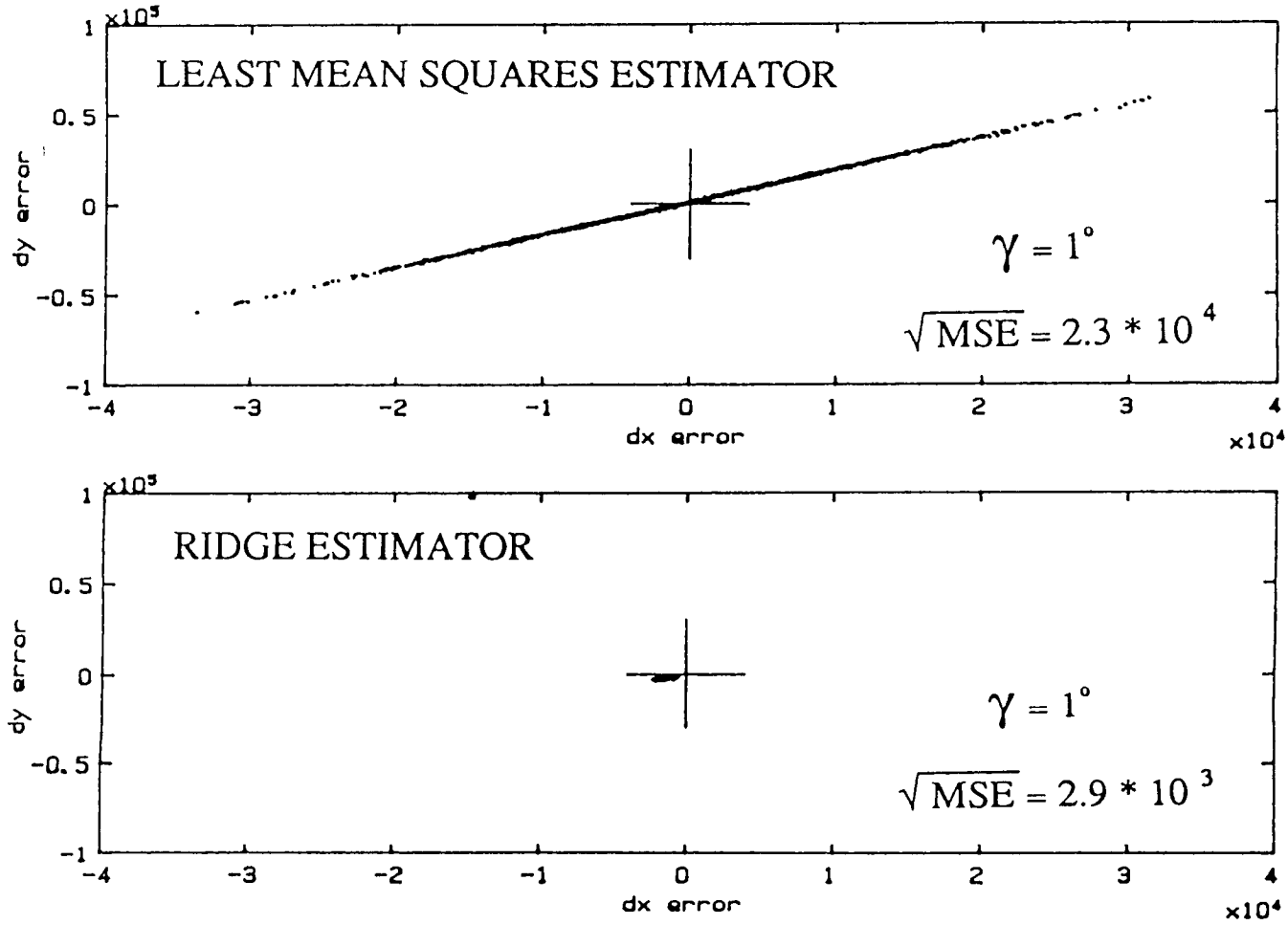


Figure 3. Comparison of the OLS estimator and the Ridge estimator through the DME simulation with the two range measurements having a small crossing angle (gamma).

OPTIMIZATION OF THE EFFECTIVE GPS DATA RATE

David S. McIntyre
Ohio University
Athens, Ohio

SUMMARY

Ohio University's Avionics Engineering Center is performing research directed towards the integration of the NAVSTAR Global Positioning System (GPS) and the Inertial Navigation System (INS) for attitude and heading determination. The integration of GPS/INS offers synergistic benefits. INS gyro drift error can be compensated by the long-term stability of GPS by means of an in-flight data monitoring algorithm. Using GPS data as a reference is more advantageous than implementing an additional INS since GPS offers a dissimilar redundancy to the attitude and heading determination configuration. In converse, the short-term stability of the INS can be used to correct or substitute for faulty GPS data due to tracking loop phase lag or data gaps because of satellite shielding.

The optimization of the effective GPS data rate is essential for the proper execution of an integrated GPS/INS in-flight algorithm. GPS attitude and heading information must be consistently available during INS outages. Present research efforts involve the development of an in-flight algorithm that maximizes the potential of integrated GPS/INS. This algorithm determines the acceptable limits of phase lag that the GPS tracking loop introduces to the flight control system (FCS) during the transmission of information. Once these calculated limits are exceeded, INS data are used to insure the continuous availability of attitude and heading information to the flight control system, as depicted in figure 1.

OVERVIEW OF GPS RECEIVER TRACKING LOOPS

Both code and carrier tracking loops are implemented in a GPS receiver to abstract navigation information. Each channel of the receiver measures an antenna phase center location on the aircraft (see figure 2). Through interferometric measurements, the receiver obtains the navigation signal. The GPS signal is tracked by the carrier and code tracking loops. Both tracking loops are based on phase-locked loop (PLL) principles. The phase-locked loop is a negative feedback system which consists of three principal components: a multiplier, a

voltage-controlled oscillator (VCO) and a loop filter. A reference signal is generated by the VCO and is continuously adjusted until its phase is equal to the phase of the carrier component of the GPS input signal.

The GPS navigation code is retrieved by an analogous method implemented in the carrier tracking loop. The input code is compared with a locally - generated reference code. An early/late detector determines whether the two codes match through the use of a correlator. Once the code is acquired, it can be combined with the carrier component to be used as the real-time reference by the carrier tracking loop. As a result, both the code and carrier tracking loop functions are mutually dependent upon one another during the continuous tracking of the GPS signal, as shown in figure 3.

DETERMINATION OF THE EFFECTIVE GPS DATA RATE

The acceptable level of phase lag introduced by the GPS guidance information into the flight control system during, for instance, a precision approach is 6 degrees for a guidance-loop bandwidth of 1.5 radians per second. If this phase lag level is exceeded, the overall phase margin of the FCS becomes too small, thereby reducing the stability of the FCS. To optimize the GPS data rate, the tracking loop bandwidth in the GPS receiver must be designed to meet the above requirement. Too narrow a bandwidth will introduce a phase lag to the flight control system since the tracking loop configuration will take longer to track the signal dynamics. In converse, too wide a loop bandwidth will introduce excessive noise into the tracking loops.

Tracking loop bandwidth optimization can be achieved by characterizing a given tracking loop circuitry in a GPS receiver through computer simulation. In a simulation environment individual component parameters can be adjusted to narrow the loop bandwidth. The effective data rate is varied by changing the loop parameters. As the data rate approaches the limiting phase lag of the flight control system, the tracking loop bandwidth parameters will approach optimum values. If no satisfactory parameters can be determined, then additional information is required, which could be provided by an INS.

A sensitivity analysis will determine which parameters have the most effect on the phase response of the tracking loop circuitry. Since numerous combinations of parameter values can be combined to produce an equivalent phase response, practical design constraints will be considered to determine the optimum set of parameter values.

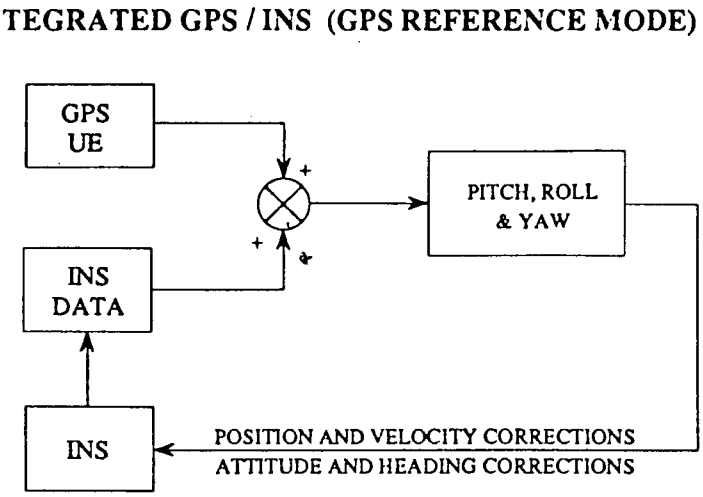


Figure 1. GPS/INS attitude and heading configuration.

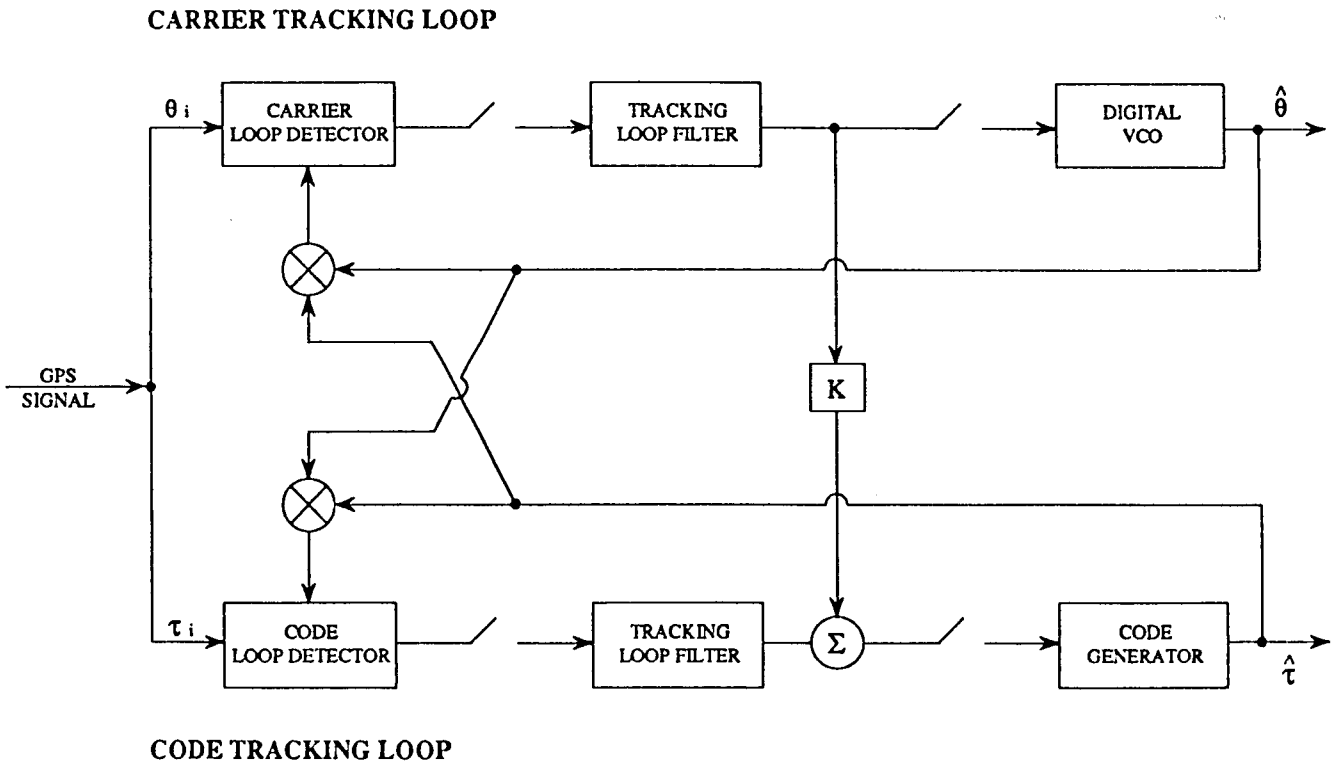


Figure 2. Simplified representation of GPS receiver tracking loops.

CONCLUSIONS

Once the effective GPS data rate is determined, the phase response of the transfer of data to the flight control system can be measured. Optimum loop parameters can be chosen to minimize phase lag through simulation.

If the effective GPS data rate is unacceptable during flight dynamics, inertial aiding must be implemented to increase the effective data rate presented to the FCS. The synergistic advantages of employing these two dissimilar systems contribute to a continuous supply of navigation information at an acceptable rate to the flight control system.

BIBLIOGRAPHY

1. Copps, E., Geier, G., Fidler, W., and Grundy, P.: Optimal Processing of GPS Signals. Special Issue on the Global Positioning System, Vol. 2, The Institute of Navigation, 1984, pp. 13-24.

2. Ward, P.: An Inside View of Pseudorange and Delta Pseudorange Measurements in a Digital Navstar GPS Receiver. Paper presented at the ITC/USA/81 International Telemetering Conference, San Diego, Ca., October 1981.

3. Kelly, R. J., and Cusick, D.: Distance Measuring Equipment and Its Evolving Role in Aviation. Advances in Electronics and Electron Physics, Vol. 68, Academic Press, Inc., 1986.

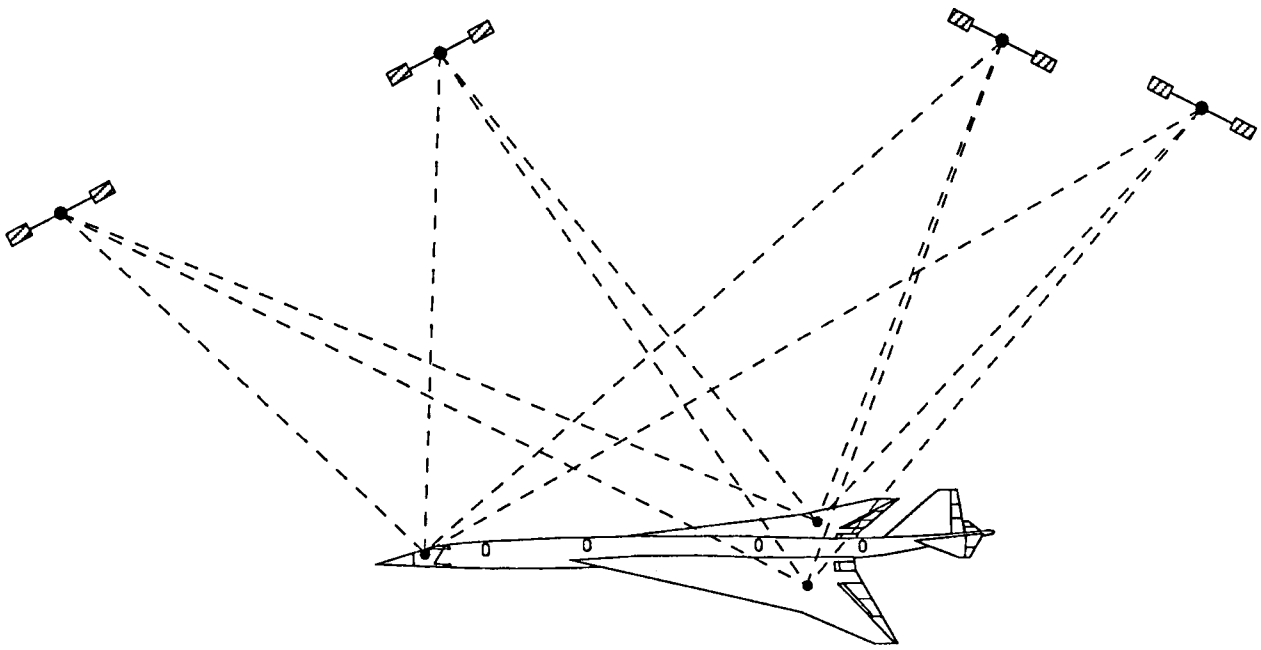


Figure 3. GPS Interferometric measurement scheme.

SOLE MEANS NAVIGATION AND INTEGRITY THROUGH HYBRID LORAN-C AND NAVSTAR GPS

Frank van Graas
Avionics Engineering Center
Ohio University

BIOGRAPHY

Ir. Frank van Graas received the B.S. degree in 1983 and the "Ingenieurs" degree in 1985, both in electrical engineering with specialization Avionics from Delft University of Technology, Delft, the Netherlands. Currently he is a research associate for the Avionics Engineering Center and Ph.D. student at Ohio University. During the past three years he has been involved with the NASA Ames differential GPS program, and system development and flight testing of GPS and LORAN-C.

ABSTRACT

A minimum of four GPS range measurements or two LORAN-C Time Differences (TDs) is normally required for a position solution for enroute navigation, area navigation, and non-precision approaches.

This paper describes a new technique that hybridizes GPS and LORAN-C used in the pseudorange mode to process efficiently all available navigation information. Emphasis is placed on combined GPS and LORAN-C timing, both for the ground/space facilities and the user.

The hybrid system has the potential to solve the GPS and LORAN-C integrity problems; more range measurements are available than required for the navigation solution.

1.0 INTRODUCTION

The NAVSTAR Global Positioning System (GPS) is expected to become operational around 1991. At that time, given the currently planned 21 satellite constellation, GPS will only qualify for a supplemental type certification. GPS does not fulfill the integrity requirements for a sole means navigation system. Several schemes have been proposed to solve the GPS integrity problem, including additional GPS satellites, geostationary satellites with ground based monitoring stations, and differential GPS. These approaches require either major government investments or significantly increased user costs (additional uplink equipment).

Another way to achieve integrity is by combining navigation systems. For the continental United States, the Long Range Navigation system, LORAN-C, combined with GPS has the potential to meet both the availability and the integrity requirements for a sole means navigation system. In addition, it is expected that the requirements for non-precision approaches will be fulfilled.

This paper is mainly concerned with the interoperability of LORAN-C and GPS. It should be emphasized that other navigation aids such as Omega, DME, IMU/INS, and altimeter could be integrated as well. The resulting navigation system should be based on a generic design that allows for effective and transparent processing of all navigation data simplifying certification and training procedures.

2.0 WHAT CONSTITUTES A SOLE MEANS NAVIGATION SYSTEM ?

Air navigation in controlled airspace requires the presence of a sole means navigation system. Although many descriptions and even a definition* exist for a sole means navigation system, not all requirements that constitute such a system are specified. Consequently, the question raised in the title of this section cannot be fully answered.

Looking at currently accepted sole means navigation systems such as VOR/DME, and considering five major performance characteristics: accuracy, availability, reliability, coverage, and integrity, both the known requirements and deficiencies of the definition for a sole means navigation system can be derived as follows: (Formal definitions for some of the requirements can be found in references [1,2]).

Accuracy: Both current accuracy requirements and future goals exist. Table 1 summarizes these accuracies for specific phases of flight in the continental United States (CONUS).

Coverage: Navigation signals must be adequate to determine position accurately within the coverage area.

Availability: This is the percentage of time that the navigation system can be used at a certain location. Availability should be close to 100 percent (VOR); the exact percentage is not specified. Generally, the requirement states that a system outage should not overload the air traffic controller. Whether an availability of 99.9% (526 minutes of outage during a year) or 99.9999% (32 seconds of outage during a year) would satisfy this requirement is not known. Also, a VOR outage only affects a relatively small area; a failing GPS satellite affects a large service area.

Reliability: This is the probability that a system will be operational continuously over a specified period of time at a certain location. A low reliability indicates that the system is likely to experience an outage over the specified period of time. Systems with long outage periods should therefore be very reliable. Solid-state VOR or DME stations have a specified Mean Time Between Failure (MTBF) of 10,000 hours [4]. For area navigation, two signals are needed at the same time. Redundant VOR/DME results then in a reliability of 95% over a period of approximately 120 days.

Integrity: This is a fairly recent requirement for navigation systems. An adequate definition of integrity used for the Institute of Navigation workshop on GPS integrity is given by [5]:

> Guaranteeing to the user (with probability p) that he will be promptly (within time T) notified when GPS system induced errors are greater than a prespecified level.

Other definitions of integrity exist, they all have the same three ingredients: a warning time T, an error limit, and a probability p. The integrity working group of RTCA Special Committee 159 has developed goals for warning times and error limits, these goals are summarized in Table 2 for navigation in CONUS [5]. A figure for probability has not been specified.

3.0 AVIONICS EQUIPMENT INTEGRATION OPTIONS

The avionics implementation of combined LORAN-C and GPS can be divided into two approaches:

1) Two Separate Systems. This approach requires a third system that combines the two receivers in one of the following ways:

   a) A (processor) system that obtains navigation data from the GPS and the LORAN-C receivers, executes the integrity checking algorithms, and provides the "best" navigation solution to the pilot (this solution may be based on data from both systems).

   b) An interface that converts the data from one system in such a way that it can be used as an extra input to the other system (e.g. LORAN-C as a pseudolite input to the GPS receiver). This approach might require minor modifications to the data receiving system.

*The 1984 Federal Radionavigation Plan [1] gives the following definition for a Sole Means Air Navigation System: An approved navigation system that can be used for specific phases of air navigation in controlled airspace without the need for any other navigation system.

Phases of Flight	Current System[1],[2] Accuracy Requirements AC 90-45A [3] 95% conf. for cross track and along track	Future System[1] Accuracy Requirements FRP-84 [1] 2 drms
Enroute Domestic	1.5 nmi	1000 m
Terminal	1.1 nmi	500 m
Non-Precision Approach	0.3 nmi	100 m

(1) System accuracy requirements do not include Flight Technical Errors.

(2) Does not include radiated signal accuracy. It is not clear from AC 90-45A how to account for these errors.

Table 1. Current and future navigation system accuracy requirements for specific phases of flight in the continental United States.

Phases of Flight	Radial Alarm Limit	Time to Alarm
Enroute Domestic	1000 m	30 s
Terminal	500 m	10 s
Non-Precision Approach	100 m	6 s

Table 2. Goals for integrity criteria as developed by the integrity working group of RTCA Special Committee 159.

2) Hybrid GPS and LORAN-C. Several grades of hybridization can be implemented. Since this system is currently non-existent, it is necessary to identify the most effective method of hybridization.

Figure 1 illustrates the separated and hybridized GPS/LORAN-C functional block diagrams. The main features of the hybrid system are the shared clock and receiver/processor assembly. The shared clock enables both systems to obtain timing information from each other, maximizing the use of all available navigation information. This will be illustrated in more detail in section 6.

Before the introduction of the navigation solution for combined GPS/LORAN-C it is necessary to take a close look at the timing of both systems. GPS maintains all timing relations between the space segment and the ground segment. However, for navigation users LORAN-C timing is established for each chain only. Current developments in LORAN-C include proposed system timing changes. This will result in a major improvement of navigational accuracies. Section 5 presents LORAN-C timing options and their effects on the navigation accuracy. LORAN-C pseudoranging is emphasized, since this allows for LORAN-C signal processing in a manner very similar to GPS, and will take advantage of the LORAN-C clock information.

4.0 GPS TIMING AND COVERAGE

The timing of the GPS system is very well defined [6]. The Master Control Station at Falcon AFS maintains GPS system time. Each satellite operates on its own reference time (space vehicle time) which is closely monitored by the GPS Ground Control Segment. Information about space vehicle clock phase offset with respect to GPS time can be calculated continuously from the navigation data transmitted by the satellites with a typical accuracy of 15 nanoseconds [7].

Currently, GPS system time is monitored to within 100 nanoseconds with respect to Universal Time, Coordinated (UTC). Offset between GPS time and UTC is also transmitted by the satellites and can be determined continuously by the user with a resolution of 1 nanosecond [6]. With the installation of three hydrogen masers at the Master Control Station, the capability exists to determine time offset between UTC and GPS time to within 30 nanoseconds [8].

The 21 satellite constellation will provide full coverage for CONUS. However, there are several hours out of each day that only four satellites are visible. During these periods, there is no ability to detect so-called "soft" GPS errors, such as might result from satellite clock degradation. Periods of limited visibility are also vulnerable to single satellite outages, causing the GPS system to be unavailable. The availability of the 21 satellite constellation is estimated to be about 75% [9].

5.0 LORAN-C TIMING AND PSEUDORANGE COVERAGE

In order to consider the effect of LORAN-C system timing on navigation and timing accuracies, four timing options with respect to LORAN-C and GPS/LORAN-C interoperability are described below. Table 3 summarizes the anticipated error budgets for pseudorange measurements to LORAN transmitters for the four timing options. Although not specifically mentioned in the following sections, pseudorange measurements are considered to be made with respect to a known time reference (UTC or GPS) at the user. In general, the knowledge of UTC or GPS time at the user is not only a function of geometry, but also depends on the magnitude of common bias errors in the pseudoranges (e.g. unmodeled propagation delays). These errors hardly affect the position solution, but will appear as an additional bias in the estimate of the system time reference (UTC/GPS). Note also that receiver hardware delays from the antenna phase center to the measurement point in the receiver are considered to be calibrated and known.

5.1 Current LORAN-C Timing.

LORAN-C Master station transmissions are synchronized to UTC within $\pm$ 2.5 microseconds. Whenever a Master drifts too far away from UTC, two methods can be used to adjust the offset: a frequency adjustment or a microphase stepper adjustment. Intentional time-steps and frequency adjustments are always announced in advance [10]. Frequency adjustments are on the same order as the drift rates of the LORAN Cesium frequency references, typically 50 - 300 nanoseconds per day.

Secondary stations are held to within $\pm$ 50 nanoseconds of the Controlling Standard Time Difference (CSTD), a reference TD established for each Master-Secondary pair, measured by a System Area Monitor (SAM). The SAM initiates Local Phase Adjustments (LPAs) at the Secondary station to maintain the CSTD. This results in an extremely stable TD for users close to the line-of-position (LOP) the SAM is located on.

The main disadvantage of the current LORAN-C timing procedure is that the time of transmission of the Secondary station varies when propagation delays to the monitor (SAM) vary. This results in an uneven error distribution with relatively large errors in areas not close to the line-of-position defined by the CSTD [11,12]. Also, propagation delay models cannot be applied easily; besides the signal path delay from the Secondary transmitter to user, the variations caused by the SAM would need to be predicted as well.

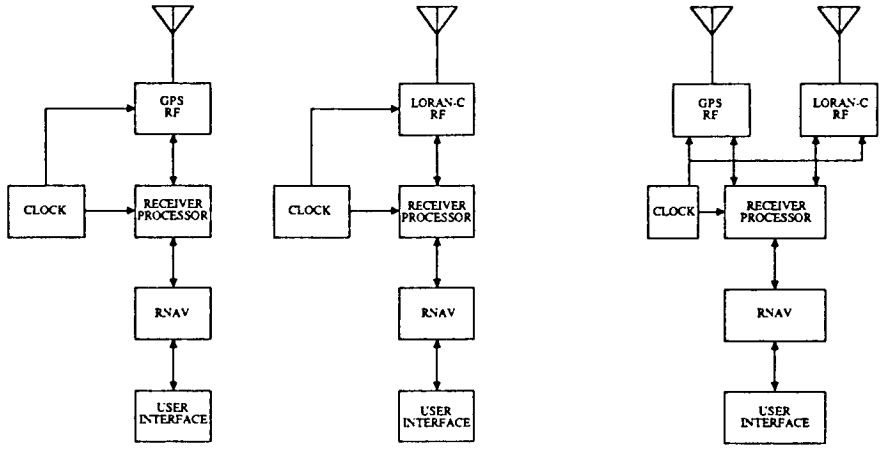
5.2 Master Station Time of Transmission Control.

Controlling the LORAN-C Master stations to an accuracy of better than 100 nanoseconds with respect to UTC will reduce the uncertainty of the LORAN-C clock phase offset relative to GPS. Including propagation uncertainties, pseudorange measurements to Master stations could be within 200 nanoseconds. Ranging to Secondaries would introduce approximately another 100 nanoseconds due to chain timing and temporal propagation effects. Users of LORAN-C only will see no net change. In fact, LORAN-C users can benefit from the increased coverage and accuracy offered by cross-chaining, approximately a factor of 2 in position accuracy with respect to current LORAN-C, mainly caused by improved geometry.

5.3 Time of Transmission Control for all LORAN-C Stations.

This option proposes a radical change in the timing control of the LORAN-C system. Each transmitter will be synchronized with respect to UTC. This approach is similar to the French direct ranging LORAN-C chain where GPS is used to monitor the timing control [12]. Time of transmission control will result in improved navigation accuracies for areas not close to line-of-positions maintained by SAMs. The main disadvantage of this option is the upward compatibility of existing LORAN-C receivers. The very high TD repeatability around the SAM will be lost, and tables for ASF corrections would have to be replaced by propagation delay models. Single chain LORAN-C users would not necessarily see an improvement in navigation accuracy when compared with Master time of transmission control. Chain timing errors and temporal propagation effects would be replaced by timing uncertainties with respect to UTC. Cross-chaining users on the other hand would benefit tremendously; all transmitters are equal, opening up a larger coverage area with good geometry. This would also increase the LORAN system availability: a failing Master station does not result in an unusable chain.

For the combined GPS/LORAN-C system, this option would be very effective: all LORAN transmitters could provide ranging with accuracies typically better than 200 nanoseconds.



(a) (b)

Figure 1. Separated (a) and hybridized (b) GPS/LORAN-C functional block diagrams.

ERROR SOURCE	LORAN-C TIMING OPTIONS			
	I	II	III	IV
TRANSMITTER - UTC SYNCHRONIZATION ERROR (ns)	±2500*	100*	100*	60*
SECONDARY-MASTER SYNCHRONIZATION ERRORS (ns)				
CHAIN TIMING	±50	±50	—	—
TEMPORAL PROPAGATION EFFECTS	0 - 50	0 - 50		
PROPAGATION ERRORS (ns) (TRANSMITTER TO USER, AFTER MODELING)	50 - 100	50 - 100	50 - 100	50 - 100
RECEIVER MEASUREMENT (ns) ERROR	25	25	25	25

I = CURRENT LORAN-C TIMING

II = MASTER STATION TIME OF TRANSMISSION CONTROL

III = TIME OF TRANSMISSION CONTROL FOR ALL LORAN-C STATION

IV = DETERMINATION OF LORAN-C TRANSMITTER OFFSETS WITH RESPECT TO GPS TIME

*NOTE THAT THE TRANSMITTER OFFSET WITH RESPECT TO UTC ONLY AFFECTS THE ESTIMATE OF UTC AT THE USER; THE POSITION SOLUTION BASED ON TRANSMITTERS FROM THE SAME CHAIN IS HARDLY AFFECTED. CROSS-CHAINING ON THE OTHER HAND IS AFFECTED SIGNIFICANTLY, UNLESS AN EXTRA TRANSMITTER IS TRACKED TO DETERMINE THE OFFSET BETWEEN DIFFERENT CHAINS.

Table 3. A comparison of error budgets for LORAN-C pseudorange measurements with respect to current and proposed system timing.

5.4 Determination of LORAN-C Transmitter Offset with respect to GPS Time.

All of the timing options discussed above should also be considered in combination with GPS receivers at each transmitter site. Data collected from GPS can be used to determine the transmitter clock offset from GPS time with an accuracy better than 30 nanoseconds. The clock offset with respect to either UTC or GPS time can then be transmitted to the user. This approach establishes, in essence, time of transmission control without affecting existing user equipment. Ranging accuracies to the transmitters will only be limited by the sum of GPS time transfer accuracy, remaining uncertainties after propagation models are applied, and user receiver errors. State-of-the-art LORAN receivers would typically achieve ranging accuracies better than 150 nanoseconds.

The use of the LORAN-C blink codes or additional pulses for the transmission of the clock offset data can be justified by the significant improvement of the timing and navigation capabilities created by this approach.

5.5 LORAN-C Pseudorange Coverage.

Use of LORAN-C in the pseudorange mode improves the geometry for users outside the center of the service area, and facilitates user clock synchronization with LORAN-C system time. If all stations are synchronized to a common time reference, coverage would significantly improve. Every combination of three stations could be used for positioning.

To predict LORAN-C pseudorange coverage for synchronized stations, a computer program was written based on a hyperbolic coverage prediction model developed for the FAA [13]. The current program predicts pseudorange coverage accounting for geometry and signal-to-noise ratio at the receiver. Figures 2 and 3 show predicted coverage for CONUS with the "mid-continent gap" filled. The mid-continent stations are indicated by triangles around the circular-shaped stations. Coverage is declared when the geometry (Horizontal Dilution of Precision HDOP), number of stations, and the signal-to-noise ratio conditions are satisfied. Figure 2 shows predicted pseudorange coverage for 3 or more stations with SNR above -10 dB and HDOP less than 4. Figure 3 shows the predicted coverage for 4 or more stations with SNR above -10 dB and HDOP less than 7.8. This figure illustrates that redundant coverage is available. Current efforts are focused on the determination of the quality of the redundant coverage. That is, given a transmitter failure, do the remaining stations satisfy the coverage requirements. One important note should be made about the atmospheric noise values used for the predicted coverage. These values are based on CCIR Report No. 322 and are rather conservative [13]. Further research is required to determine atmospheric noise values that are in closer agreement with actual measurements.

From Figure 2 and Table 3 it can be concluded that positioning with synchronized LORAN-C stations can result in navigational accuracies on the order of 100 - 200 meters throughout CONUS.

6.0 AIRBORNE NAVIGATION SOLUTION

Several schemes can be implemented to combine the navigation data from LORAN-C and GPS. For example, a humongous Kalman filter could be developed that processes all available GPS pseudorange measurements and LORAN-C time differences. Even though such an approach promises to be optimal, certification procedures are most likely to be hindered by the physically impossible task to ensure the performance of the navigation filter under all input conditions [14]. Another concern is the complexity of modifications caused by the addition or deletion of navigation sensors or upgrades of existing sensors. These modifications should not necessitate a new system design with related certification and training procedures. Instead, the system should recognize the change and take appropriate actions.

Therefore, the navigation solution should be based on a generic design that emphasizes effective, modular, and transparent rather than optimal processing.

A system design philosophy that satisfies the above requirements could be based on the conversion of all sensor inputs into comparable quantities. Differences in sensor performance can be accounted for by assigning weights to the individual sensor measurements. These weights can for instance be determined by the magnitude and variance of measurement residuals, differences between actual measurements and predicted measurements based on previous data.

Pseudorange measurements are in common to both GPS and LORAN-C. The main advantages of using pseudoranges over time differences are the additional clock phase offset information and the option to use single transmitters instead of pairs. For instance, 4 GPS satellites could be used for navigation in combination with only one LORAN station to achieve integrity.

Noise on the pseudorange measurements can be effectively reduced by range domain filtering techniques [15,16]. This eliminates the possibility of navigation domain filtering divergence and allows for straightforward filter tuning. Although process noise cross-correlation terms are discarded in the range filters it was shown for stand-alone GPS that the overall system performance is essentially that of navigation domain filters [16]. Similar results may be expected for a solution based on both GPS and LORAN-C pseudorange measurements.

6.1 GPS Pseudorange Measurements.

Pseudorange measurements to GPS satellites are made by taking the difference between the measured time of signal arrival and the known time of signal transmission, corrected for known and estimated error sources. Figure 4 illustrates the ranging geometry. The general equation for the measured pseudorange is given by:

$$P_i(t) = |\vec{S}_i(t - \beta_i(t)) - \vec{u}(t)| + c_{GPS}(T_{GPS}(t) - T_{S_i}(t) + d_{GPS_i}(t,r)) \qquad (1)$$

where:

$\vec{S}_i$	position vector satellite i
β_i	line-of-sight travel time for signals from satellite i
$\vec{u}$	user position vector
c_{GPS}	GPS value for the speed of light
T_{GPS}	user clock offset from GPS system time
T_{S_i}	satellite i clock offset from GPS system time
d_{GPS_i}	delay for measurement i caused by GPS error sources

Typical GPS pseudorange accuracies are on the order of 10-40 meters (C/A code), mostly depending on ephemeris uncertainty, ionospheric and tropospheric delays, and intentional signal degradation, if active. GPS position accuracy is specified to be 100 meters 2 drms.

6.2 LORAN-C Pseudorange Measurements.

LORAN-C pseudorange measurement geometry is different from GPS; in the coverage area, LORAN-C ground waves basically travel great-circle distances. A receiver at sea-level will interpret the signals as if they came from transmitters located in the locally level plane at distances equal to great-circle distances to the transmitters, as depicted in Figure 5. The LORAN-C

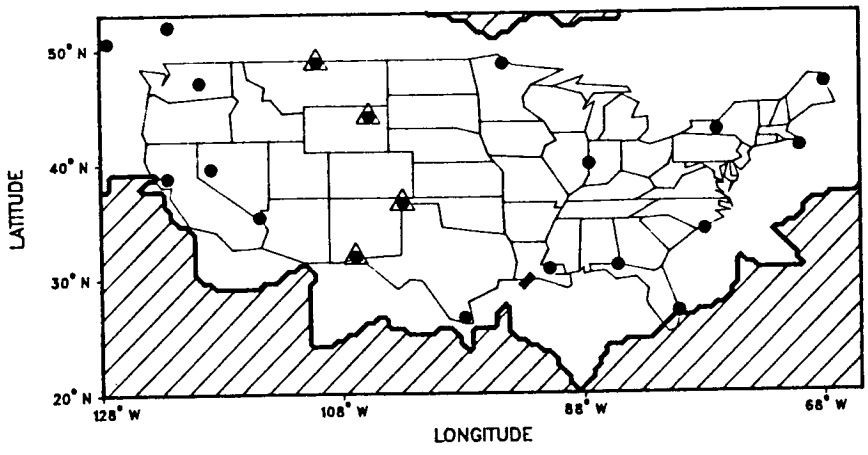


Figure 2. Preliminary LORAN-C predicted pseudorange coverage with the "mid-continent gap" filled. Coverage is computed under the following assumptions:

- All stations are synchronized
- All-in-view solution using 3 or more stations
- SNR greater than -10 dB
- HDOP less than 4
- Receiver bandwidth = 20 kHz
- Atmospheric noise values used are for the summer season, based on CCIR Report No. 322
- Search increment = 0.5°

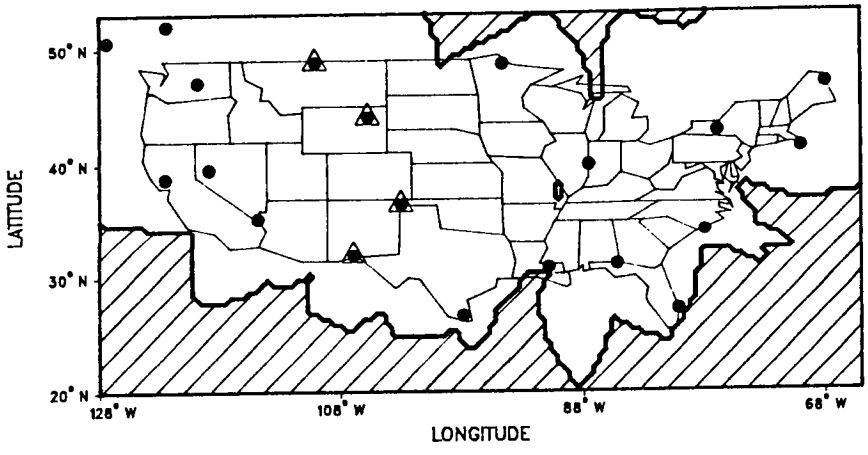
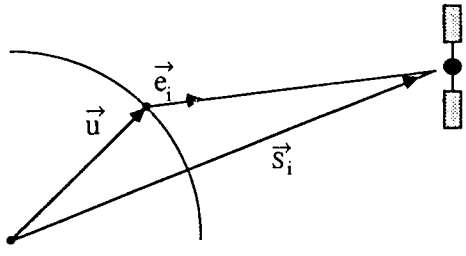


Figure 3. Preliminary LORAN-C predicted pseudorange coverage with the "mid-continent gap" filled. Coverage is computed under the following assumptions:

- All stations are synchronized
- All-in-view solution using 4 or more stations
- SNR greater than -10 dB
- HDOP less than 7.8
- Receiver bandwidth = 20 kHz
- Atmospheric noise values used are for the summer season, based on CCIR Report No. 322
- Search increment = 0.5°

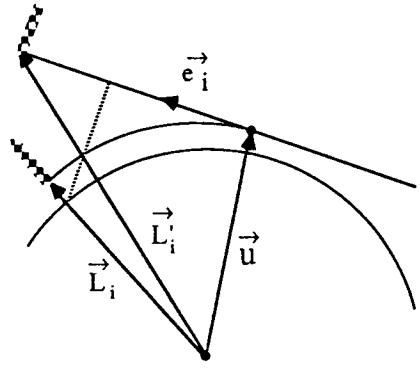


$\vec{S}_i$ position vector satellite i

$\vec{u}$ user position vector

$\vec{e}_i$ line-of-sight vector for satellite i

Figure 4. GPS ranging geometry.



$\vec{L}'_i$ position vector LORAN-C transmitter corrected for earth curvature

$\vec{L}_i$ position vector LORAN-C transmitter

$\vec{u}$ user position vector

$\vec{e}_i$ line-of-sight vector for transmitter i

Figure 5. LORAN-C ranging geometry.

pseudorange equation is given by:

$$P_i(t) = |\vec{L'_i}(t) - \vec{u}(t)| + c\,(T_{LC}(t) - T_{L_i}(t) + d_{LC_i}(t,r)) \qquad (2)$$

where:

$\vec{L'_i}$	position vector LORAN-C transmitter corrected for earth curvature
$\vec{u}$	user position vector
c	speed of light in vacuum
T_{LC}	user clock offset from LORAN-C time
T_{L_i}	transmitter i clock offset from LORAN-C time
d_{LC_i}	delay for measurement i caused by LORAN-C error sources

LORAN-C pseudorange performance is given in Table 3 for different timing options. Stand-alone LORAN-C positioning using pseudoranges yields typical accuracies of 100 - 200 meters for HDOP less than 4.

6.3 Hybrid GPS/LORAN-C Navigation Solution.

From the pseudorange geometries given in Figures 4 and 5, and equations (1) and (2), the user position and clock biases can be obtained by solving the following set of equations:

$$\left[\begin{array}{ccc|cc} e_{11} & e_{12} & e_{13} & 1 & 0 \\ e_{21} & e_{22} & e_{23} & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ e_{n1} & e_{n2} & e_{n3} & 1 & 0 \\ \hline e_{(n+1)1} & e_{(n+1)2} & e_{(n+1)3} & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ e_{(n+m)1} & e_{(n+m)2} & e_{(n+m)3} & 0 & 1 \end{array}\right] \left[\begin{array}{c} X \\ Y \\ Z \\ \hline T_{GPS} \\ T_{LC} \end{array}\right] = \left[\begin{array}{c} \vec{e_1} \cdot \vec{S_1} - P_1 + T_{S_1} \\ \vdots \\ \vec{e_n} \cdot \vec{S_n} - P_n + T_{S_n} \\ \hline \vec{e_{n+1}} \cdot \vec{L'_{n+1}} - P_{n+1} + T_{L_{n+1}} \\ \vdots \\ \vec{e_{n+m}} \cdot \vec{L'_{n+m}} - P_{n+m} + T_{L_{n+m}} \end{array}\right] \qquad (3)$$

Variations on these equations result when different LORAN-C timing options are considered. As an example, option IV, "Determination of LORAN-C transmitter offset with respect to GPS time," would remove the user clock offset uncertainty with respect to LORAN-C system time from the set of unknowns. In other words, only 4 measurements are needed to solve for user position and clock offset from GPS time.

Generally, a total of at least 8 measurements will be available throughout CONUS, leaving 3 or 4 measurements (depending on accuracy requirements) for integrity checking. Current LORAN-C timing would either require some of the redundant measurements to solve for clock offsets between chains and/or stations, or GPS could be used to (continuously) calibrate LORAN measurements. In the latter case, calibration could start with 5 GPS satellites, or only 4 GPS satellites and either one accurate LORAN measurement, or one additional measurement such as (airport) altitude.

6.4 Hybrid GPS/LORAN-C Integrity.

As indicated in the previous sections, integrity can be obtained through utilization of redundant pseudorange measurements from GPS and LORAN-C. Recently, several papers have been published on the subject of user autonomous integrity checking with redundant measurements [17-20].

Measurement residuals are commonly used to indicate a bad signal. This may be accomplished through maximum likelihood detection schemes, a parallel bank of Kalman filter failure hypothesis testers, or even through direct comparison of the residuals. Further research is needed in this area to determine the most effective method, once the proper requirements for integrity are established.

7.0 EXPERIMENTAL HYBRID SYSTEM DESIGN

Early in-flight comparison of GPS and LORAN-C indicated 2-dimensional differences of up to 300 meters for data collected across southern and central Ohio [21]. Most of the navigation error was found to be inherent to the LORAN-C part of the system. A TI 9900 LORAN receiver and the Experimental Dual Channel GPS receiver were used during these tests [22]

Currently, work is ongoing to replace the hyperbolic LORAN receiver with a RACAL Megapulse Accufix 500 LORAN-C receiver. The resulting system will be configured as depicted in Figure 1b. Both the GPS and the LORAN-C receiver can be fully controlled by an external computer system. The main advantage is that tracking and data smoothing filters can be adjusted. This eliminates large navigation errors during maneuvers due to filter lag which is especially the case for most LORAN-C receivers. All data from the receivers will be collected on magnetic tape and disk for postprocessing on the ground in combination with ground based tracker data. This will create a data base for performance evaluation of navigation and integrity algorithms.

8.0 CONCLUSIONS AND RECOMMENDATIONS

A sole means navigation system does not only call for integrity, but also for coverage, reliability, availability and accuracy. Even though ground monitored GPS will provide integrity, availability is still not sufficient. One satellite outage can affect a large service area for several hours per day. The same holds for differential GPS, a total satellite outage cannot be corrected for. To obtain sufficient coverage, extra measurements are needed. Either in the form of extra GPS satellites (expensive) or through redundant measurements from other systems. LORAN-C is available and will, hybridized with GPS, result in a system that has the potential to satisfy the requirements for a sole means navigation system for use in the continental United States.

Assumptions are made about the qualification sole means, mainly based on current sole means systems such as VOR/DME. In order to allow for system design that will

satisfy sole means requirements, it is recommended that a definition of a sole means navigation system be established. This definition must include requirements for availability, reliability, and integrity currently not specified.

In addition to the definition of a sole means navigation system, certification requirements must be established for hybrid navigation systems. This will allow for design and production of a new generation of airborne navigation systems that will reduce overall system costs and simplify training procedures.

The current LORAN-C navigation and timing system could be greatly enhanced by upgrading the synchronization between stations. It is recommended to implement time of transmission control for all LORAN-C stations under the condition that the impact on current users will be minimal. Otherwise, GPS receivers should be installed at each LORAN-C station to determine the station clock offset with respect to GPS. These offsets should be transmitted to the users using blink codes or additional pulses. This would establish, in essence, time of transmission control without affecting current users. Either timing option would significantly increase the LORAN-C coverage area (see Figures 2 and 3) and also result in navigational accuracies on the order of 100 - 200 meters throughout CONUS. In addition, hybrid GPS/LORAN-C would have a minimum of three redundant measurements to insure system availability and integrity.

ACKNOWLEDGEMENTS

This work is supported by the Federal Aviation Administration and NASA Langley Research Center under the Joint University Program in Air Transportation Systems (Grant NGR-009-017). The author is indebted to Dr. Richard H. McFarland, Dr. Robert W. Lilley, and Mr. William L. Polhemus for their suggestions and remarks.

REFERENCES

[1] Federal Radionavigation Plan - 1984. DOD-4650.4 and DOT-TSC-RSPA-84-8, U.S. Department of Defense and U.S. Department of Transportation, Washington, D.C., 1984.

[2] Braff, R., Shively, C.A., Zeltser, M.J., "Radionavigation System Integrity and Reliability," Proceedings of the IEEE, Vol. 71, No. 10, October 1983.

[3] Department of Transportation Federal Aviation Administration Advisory Circular No. AC 90-45A: "Approval of Area Navigation Systems for Use in the U.S. National Airspace System," February 1975.

[4] Blythe, P.D., "Reliability of Navigation Systems," Draft Report, ARINC Research Corporation, Contract No. DTFA01-80-C-10030, March, 1982.

[5] GPS Integrity Workshop held at Interstate Electronics Corporation in conjunction with ION Technical Meeting, Anaheim, CA, 23 January 1987.

[6] Rockwell International Corporation, "ICD-GPS-200," 26 September, 1984.

[7] Bowen, R., Swanson, P.L., Winn, F.B., Rhodus, N.W., and Feess, W.A., "Global Positioning System Operational Control System Accuracies," Navigation: Journal of the Institute of Navigation, Vol. 32, No. 2, Summer 1985.

[8] Green, G.B., "Global Positioning System A Status Report," Proceedings of the Fourth National Technical Meeting of the Institute of Navigation, Anaheim, CA, January 20-23, 1987.

[9] Jorgensen, P.S., "Achieving GPS Integrity and Eliminating Areas of Degraded Performance," Proceedings of the Forty-Third Annual Meeting of the ION, Dayton, Ohio, June 23-25, 1987.

[10] U.S. Naval Observatory, "Daily Time Differences and Relative Phase Values", Washington, D.C.

[11] Campbell, L.W., Doherty, R.H., and Johler, J.R., "LORAN-C System Dynamic Model Temporal Propagation Variation Study," Report No. DOT-CG-D57-79, July 1979.

[12] Vicksell, F.B., and Goddard, R.B., "Implementation and Performance of the TOT Controlled French LORAN Chain," Proceedings of the 15th Annual Technical Symposium of the Wild Goose Association, New Orleans, Louisiana, 21-24 October, 1986.

[13] El-Arini, M.B., "Airport Screening Model for Nonprecision Approaches Using LORAN-C Navigation," Contract No. DTFA01-84-C-00001, MITRE report No. MTR-83W180, May 1984.

[14] Kruczynski, L.R., and Eschenbach, R.F., "New Data on Two Satellite Algorithms Without an Atomic Clock," Proceedings of the Fourth National Technical Meeting of the Institute of Navigation, Anaheim, CA, 20-23 January, 1987.

[15] Paielli, R.A., "Range Filtering for Sequential GPS Receivers," Proceedings of the Fourth National Technical Meeting of the ION, Anaheim, CA, 20-23 January, 1987.

[16] Van Graas, F., "Discrete Filtering Techniques Applied to Sequential GPS Range Measurements," NASA Contract NAS 2-11969, NASA-CR-177472, 1987.

[17] Kerr, T., "Decentralized Filtering and Redundancy Management/Failure Detection for Multisensor Integrated Navigation Systems," Proceedings of the Second National Technical Meeting of the ION, San Diego, CA, January 15-17, 1985.

[18] Brown, R.G., and Hwang, Y.C., "GPS Failure Detection by Autonomous Means Within the Cockpit," Proceedings of the Forty-Second Annual Meeting of the ION, Seattle, Washington, June 24-26, 1986.

[19] Lee, Y.C., "Analysis of Range and Position Comparison Methods as a Means to Provide GPS Integrity in the User Receiver," Proceedings of the Forty-Second Annual Meeting of the ION, Seattle, Washington, June 24-26, 1986.

[20] Parkinson, B.W., Axelrad, P., "Simplified GPS Integrity Checking with Multiple Satellites," Proceedings of the Forty-Third Annual Meeting of the ION, Dayton, Ohio, June 23-25, 1987.

[21] Morrell, F.R., Compiler, "Joint University Program for Air Transportation Research - 1986," Proceedings of a conference held at NASA Langley Research Center, Hampton, Virginia, January 8-9, 1987, NASA CP-2502.

[22] Campbell, S.D., and LaFrey, R.R., "An Experimental GPS Navigation Receiver for General Aviation: Design and Measured Performance," Contract No. DOT-FA79-WAI-091, Report No. FAA-RD-83/26, September 27, 1983.

WEATHER DATA DISSEMINATION TO AIRCRAFT

Richard H. McFarland, Ph.D.
Craig B. Parker
Avionics Engineering Center
Ohio University
Athens, Ohio

Abstract

Documentation exists that shows weather to be responsible for approximately 40 percent of all general aviation accidents with fatalities. Weather data products available on the ground are becoming more sophisticated and greater in number. Although many of these data are critical to aircraft safety, they currently must be transmitted verbally to the aircraft. This process is labor intensive and provides a low rate of information transfer. Consequently, the pilot is often forced to make life-critical decisions based on incomplete and outdated information.

Automated transmission of weather data from the ground to the aircraft can provide the aircrew with accurate data in near-real time. The current National Airspace System Plan calls for such an uplink capability to be provided by the Mode S Beacon System data link. Although this system has a very advanced data link capability, it will not be capable of providing adequate weather data to all airspace users in its planned configuration. This paper delineates some of the important weather data uplink system requirements, and describes a system which is capable of meeting these requirements. The proposed system utilizes a run-length coding technique for image data compression and a hybrid phase and amplitude modulation technique for the transmission of both voice and weather data on existing aeronautical Very High Frequency (VHF) voice communication channels.

Background

The National Transportation Safety Board (NTSB) broad cause/factor assignments for general aviation accidents in the United States show weather to be the overall cause of approximately 40 percent of the general aviation accidents with fatalities occurring from 1980 to 1984 [1]. According to these data, only pilot error outranks weather as the most frequent overall cause of fatal accidents in general aviation. Similarly, the data for United States air carriers show weather to be the overall cause of approximately 35 percent of the air carrier accidents with fatalities occurring in the same time period [2].

Not included in the above data is the August 1985 Delta L-1011 crash in which 135 persons lost their lives when the pilot attempted to penetrate a severe thunderstorm containing a microburst on final approach at the Dallas/Fort Worth International Airport. This crash led the NTSB to express concern that although the Federal Aviation Administration (FAA) had addressed nearly all of the actions proposed by the Safety Board since 1973, one important problem was not adequately being addressed. In the aircraft accident report for the Delta disaster, the NTSB identifies that problem as "the communication of hazardous weather information available from ground sensors to the flightcrew in time for the information to be useful in go/no-go decision making." The NTSB continues: "Current procedures to relay NWS [National Weather Service] information through the ATC [Air Traffic Control] system are not and will never be adequate for dynamic weather conditions" [3].

The aircraft pilot is the principal decision maker when in flight, and has the ultimate responsibility for the safety of the aircraft. Unfortunately, the information the pilot needs to make informed decisions about navigation in and around areas of severe weather simply is not available in today's National Airspace System.

Weather information is provided to users of the National Airspace System by a weather system which has been developed and is operated through a joint venture involving the FAA, the Department of Defense (DOD), the National Oceanic and Atmospheric Administration (NOAA), and the civil aviation community. Many of the weather data products utilized in the system are provided by the National Weather Service (NWS). The system is known as the Aviation Weather System, and is represented functionally in figure 1.

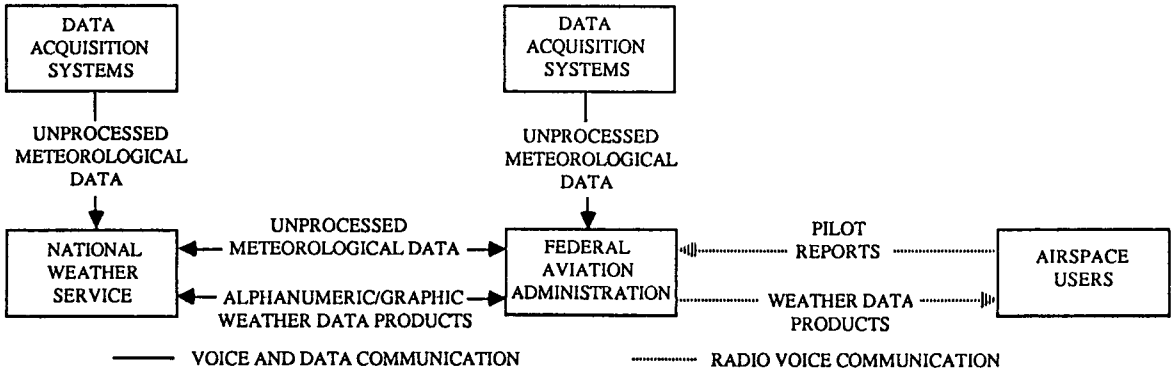


FIGURE 1 FUNCTIONAL DIAGRAM OF AVIATION WEATHER SYSTEM

In spite of significant advances in the ability to gather and disseminate weather data on the ground, the Aviation Weather System continues to rely upon voice communication for the dissemination of weather data to the aircraft. As a result, the system is severely limited in its ability to disseminate current weather data which is operationally significant to the individual end user.

The transmission of current weather data products from the ground to the aircraft by data uplink is widely recognized as a feasible solution to the weather data dissemination problem. The FAA has the responsibility to determine what meteorological services are required for aircraft efficiency and safety, and has committed to providing this service. The Aviation Weather System Plan states that at some point in the future the pilot will have in-flight access to required weather data "automatically and by request via data link" [4].

Requirements

Weather Data

Airspace user organizations and several U.S. Government agencies have expressed recommendations for the improvement of the Aviation Weather System. These recommendations are included in the Aviation Weather System Plan [5]. Recommendations with specific applicability to the uplink of weather data to aircraft are included in table 1.

Research in this area has resulted in the identification of certain weather data products which would improve aircraft safety and efficiency were they transmitted to the pilot by real time data uplink [6]-[8]. These weather data products are as follows:

1. Hazardous weather conditions
2. Radar precipitation reflectivity patterns
3. Surface analysis (SA)
4. Terminal and area forecast (FT and FA)
5. Critical weather maps
6. Text information:
   - Pilot Reports (PIREPS)
   - Significant Meteorological Information (SIGMETS)
   - Airman's Meteorological Information (AIRMETS)
   - Notices to Airmen (NOTAMS)
7. Satellite images

Organization	Expressed User Needs
ALPA	* NWS weather radar display
AOPA	* Timely critical forecasts * Self-briefing capability * Direct access to weather data
ATA	* Satellite images * NWS/FAA weather radar display
NBAA	* Pre-flight and in-flight weather data/briefings * Mass dissemination of Aviation Weather System data
NTSB	* Real time display and classification of precipitation and turbulence
NOAA/FAA/ NASA[1]	* Automated transmission of hazardous weather areas

ALPA - Air Line Pilots Association
AOPA - Aircraft Owners and Pilots Association
ATA - Air Transport Association
NBAA - National Business Aircraft Association
NASA - National Aeronautics and Space Administration
[1]1981 jointly sponsored workshop on meteorological and environmental inputs to aviation systems

TABLE 1
RECOMMENDED WEATHER DATA PRODUCTS
FOR UPLINK TO AIRCRAFT [5]

Microburst detection systems such as the Low Level Wind Shear Alert System (LLWAS) have been shown by simulation to be as much as 94 percent effective in the avoidance of microbursts if displayed graphically to the pilot. This compares to 43 percent if warnings are transmitted only by voice [9]. For such systems to provide the level of safety of which they are capable, the pilot needs to have the data available in graphical form.

Based on these facts, it is apparent that any system developed for the uplink of weather data will be required to provide some type of graphics transmission capability. Some of the more important graphics weather data would consist of real time radar precipitation reflectivity patterns and graphical microburst information. Such data would provide the pilot useful and up-to-date information about areas of operationally significant weather, and most importantly, areas of hazardous weather.

Many larger aircraft are equipped with on-board weather radar systems which provide similar information as do ground-based radar systems. These systems do not, however, have the capability of ground-based systems. Due to equipment size, cost, and weight considerations, airborne systems operate at X-Band and at lower power levels than ground-based systems. X-Band energy is more readily absorbed by precipitation and is also attenuated more in free-space than the S-Band and C-Band energy which is utilized by many ground-based systems. For these reasons, airborne systems have lesser range and are subject to greater precipitation shadowing than their ground-based counterparts. Additionally, airborne systems do not provide 360 degree azimuthal coverage. Although airborne radar systems can provide very useful precipitation reflectivity information, they cannot provide as accurate and detailed information as ground-based radar systems.

Ground-based weather radar data are currently available to the Center Weather Service Unit (CWSU) within the Aviation Weather System. Improved radar systems such as the Next Generation Weather Radar (NEXRAD) and the Terminal Doppler Weather Radar (TDWR) will provide additional information such as doppler echoes for location of windshear. The uplink of such information is certainly technically feasible and would greatly enhance the safety of the National Airspace.

Data Transmission

Transmission of weather data to the aircraft must be accomplished in such a way as to minimize the use of an already overcrowded frequency spectrum, while at the same time provide an adequate temporal update rate for near-real time weather data dissemination. The pilot or aircrew should be able to access these weather data without expending a great amount of additional effort since workloads are already quite heavy. Additionally, the transmission mode should provide adequate spatial coverage to allow uplink of weather data at any flight altitude and especially during final approach and departure. Coverage on the ground at the terminal would be very useful for updated pilot weather briefings after departure delays. Finally, the transmission of weather data products should be a service provided by the FAA as part of the Aviation Weather System. In this way airborne equipment requirements can be standardized and transmission can be accomplished using existing aeronautical radio frequency channels without competition for transmission rights from commercial interests. As a result, quality weather data products will be available to all airspace users throughout the United States. The transmission requirements identified above are summarized as follows:

1. Near-real time transmission with spectrum conservation

2. Ease of pilot access to weather data

3. Adequate spatial coverage

4. Weather uplink service provided by FAA

An ideal weather uplink system implementation would meet all of the described requirements with a minimal additional investment on the part of the aircraft operator and with minimal equipment, facilities, and maintenance cost to the FAA.

Mode S Data Link

The Mode Select (Mode S) Beacon System is the planned replacement for the current Air Traffic Control Radar Beacon System (ATCRBS) [10],[11]. The Mode S system has evolved from earlier development work on the Discrete Address Beacon System (DABS). The system is capable of providing radar surveillance with selective interrogation of aircraft, and also of ground-to-air and air-to-ground data exchange. The current Aviation Weather System Plan relies upon this data exchange capability for the uplink of ground-based weather data products [12]. For the reasons described in the following discussion, this system is not capable of providing adequate weather data uplink services to all users of the National Airspace.

The Mode S interrogator is capable of transmitting three classes of messages to the aircraft. These are as follows:

1. Surveillance data

2. Standard message

3. Extended-length message (ELM)

These messages consist of a 56-bit or 112-bit data block. Included in the data block is a 24-bit discrete address overlaid with parity check bits which provides for selective interrogation of aircraft and also for error detection. The messages are transmitted at a 4 Mbit/s data rate using differential phase shift keying (DPSK). Effective utilization of the parity check bits will allow for an overall undetected bit error rate of less than 10^{-7}. The interrogator is capable of servicing 700 users within the service volume, and so the overall data exchange rate is very high.

The standard message and the ELM are capable of providing general purpose data transmission. The standard message provides surveillance data along with a 56-bit data link message. This message type is intended primarily for messages not requiring large numbers of consecutive bits, and replaces the surveillance message while still providing the surveillance data. The ELM does not contain the surveillance data, and thus cannot substitute for the surveillance message. It is designed to accommodate transfer of large amounts of data, and consists of up to 16 112-bit data blocks of which 80-bits are available for the message. This results in one message block of up to 1280 bits being transmitted per ELM. Because graphic weather data products will consist of relatively large amounts of data, the ELM is the most efficient message format for the uplink of this information.

The ELM data rate available to the individual user is limited by the ability of the interrogator to service all of the users within the service volume. For terminal interrogators, the antenna scan time is approximately 4 seconds. Allowing one 16-segment ELM for an individual user in any one antenna scan would limit the individual user to 1280 bits in the 4 second scan time. The Minimum Operational Performance Standards (MOPS) for the Mode S airborne equipment require the aircraft transponder data link interface to be capable of handling this same amount of data in the same 4 second period of time [13]. Both performance criteria result in an overall effective ELM data rate of 320 bits/s for the individual user. This number is considerably less than the overall system data handling capability because the individual user is being communicated with selectively.

The Mode S selective communication capability is very important for the transfer of data used for air traffic control and would be useful for the transmission of user-specific weather data while en-route. It is not required, however, for the transmission of weather data products that affect all airspace users in a particular terminal area. In fact, the data rate limitation of a fully loaded Mode S interrogator in the terminal area could slow the transmission time of graphic weather data products at a time when access to current weather data is extremely important.

One other very important Mode S data link limitation for the uplink of weather data is the system spatial coverage. The National Airspace System Plan calls for the first 137 Mode S systems to provide coverage down to 12500 feet above mean sea level (msl) and to the ground at high density terminals. These systems are to be in place in the early 1990's. The second contract for an

additional 60 systems would then provide coverage down to 6000 feet msl by 1993 [14]. As a result, any airspace user at altitudes of less than 12500 feet msl would not have weather uplink capability unless in the vicinity of a high density terminal equipped with Mode S surveillance. After 1993, this coverage would be lowered to 6000 feet msl, but would still not provide for approaches and departures from any terminal not equipped with Mode S surveillance. This is not a significant problem for the air carriers, but leaves general aviation aircraft without weather uplink capability for approaches and departures from the terminals most often used by these aircraft.

For these reasons, the Mode S data link alone cannot meet the weather data uplink requirements for all airspace users. Although the Mode S data link capability is important for the transmission of air traffic control data and en route weather data, supplemental systems will be necessary at lower density terminals and would greatly enhance the individual user data rate capability at higher density terminals.

VHF Voice/Data Transmission

The continuous broadcast of weather data by VHF data uplink at the terminal area would meet the data uplink requirements and could be accomplished simultaneously with existing voice communications at these locations. The resulting system would not require any additional spectrum, would support a high data rate regardless of number of users, and would provide line-of-sight (LOS) coverage down to ground level. Additionally, the system could be implemented with only modifications to existing ground-based and airborne VHF communications equipment.

Voice and digital data can be transmitted simultaneously through the use of two independent modulations of the same carrier. Continuous wave radio communication is accomplished through the modulation of the amplitude, phase, or frequency of a continuous wave radio frequency carrier. The current VHF aeronautical radio frequency channels used for communication between the ground and the aircraft utilize amplitude modulation (AM) for the transmission of voice. In an amplitude modulated continuous wave, the phase and frequency carry no information. If phase modulation (PM) of this same carrier is accomplished by a digital data stream, then both voice and data can by transmitted simultaneously on the same carrier.

An amplitude modulated carrier can be expressed as:

$$A_c(1 + k_v m(t))\cos(2\pi f_c t) \qquad (1)$$

where A_c is an arbitrary constant, k_v is the amplitude sensitivity of the modulator, m(t) is the modulating waveform, and f_c is the carrier frequency. Similarly, a phase modulated carrier can be expressed as:

$$A_c\cos(2\pi f_c t + \theta(t)) \qquad (2)$$

where $\theta(t)$ is the phase modulation resulting from the modulating waveform. Combining amplitude and phase modulations yields the hybrid modulation waveform:

$$A_c(1 + k_v m(t))\cos(2\pi f_c t + \theta(t)) \qquad (3)$$

By letting m(t) be the voice modulation, and $\theta(t)$ be the phase modulation resulting from a digital data stream of weather information, simultaneous communication of these data can be accomplished.

The hybrid modulated carrier can be generated and received as shown in figure 2. The described modulation technique works perfectly provided that the signal is not subject to non-linear processes such as amplitude or bandwidth limiting. Under these ideal conditions, the AM envelope detector only detects the amplitude modulated voice information. Similarly, an ideal phase detector responds only to zero crossings and is not sensitive to amplitude modulation. Thus only the phase modulated data information is detected by the phase detector. Of course, the phase information is only retrievable if the AM modulation index (μ) of the process is limited to values of less than 1:

$$|\mu| = |k_v m(t)| < 1, \quad \text{for all } t \qquad (4)$$

It is equally important to maintain limits on μ in existing AM systems in order to prevent overmodulation while still providing adequate sideband power levels for efficient information transmission. Specifications for ground-based aeronautical VHF transmitters require μ to be greater than 0.7 but not to exceed 1.0 [15].

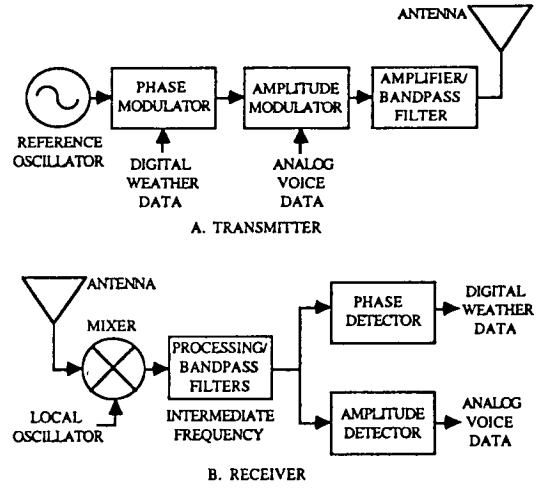


FIGURE 2
TRANSMITTER (A) AND RECEIVER (B) FOR SIMULTANEOUS TRANSMISSION OF VOICE AND DATA

Limitations

Unfortunately, an actual radio communication system cannot achieve ideal linear conditions. Amplitude limiting occurs in modern Class C output amplifiers, and would be necessary in the receiver in order to limit the dynamic range of signals presented to the phase detector. Furthermore, filtering performs bandwidth limiting at the transmitter power amplifier output for harmonic attenuation and frequency spectrum control, and in the receiver for selectivity.

These limitations affect the performance of both the voice and data transmission. Voice transmission performance is degraded by amplitude fluctuations caused when bandlimiting removes high frequency energy from the carrier. This high frequency energy is created by the phase modulation of the carrier. Data transmission performance is affected by both bandwidth and amplitude limiting. Amplitude limiting results in a lower effective signal level present at the phase detector, and also creates an interphasor crosstalk mechanism for the digital data. Additionally, bandwidth limiting creates intersymbol interference for the digital data.

Spectrally Efficient Phase Modulations

Before analyzing the effects of bandwidth and amplitude limiting on system performance, it is necessary to consider the specific type of phase modulation utilized. An overview of two spectrally efficient phase modulation techniques is presented here [16],[17].

Quadrature phase shift keying (QPSK) and minimum shift keying (MSK) have both received increasing attention in recent years due to the common characteristic of transmitting maximum amounts of information while occupying relatively small amounts of frequency spectrum. Both modulation techniques accomplish this by utilizing one of four possible phase states for the representation of data. Using this technique, each phase state corresponds to one symbol composed of two data bits in the modulating data stream.

Quadrature Phase Shift Keying (QPSK)

The QPSK signal waveform can be expressed as:

$$g(t) = A_s\cos[2\pi f_c t + \tfrac{1}{4}(2i - 1)\pi] \qquad (5)$$

where A_s is an arbitrary constant, i= 1,2,3, or 4 depending on the symbol transmitted by the data stream during the symbol period T, and f_c is the carrier frequency chosen to be an integer multiple of 1/T. This signal is more easily analyzed in the quadrature form:

$$g(t) = g_i(t)\cos(2\pi f_c t) - g_q(t)\sin(2\pi f_c t) \qquad (6)$$

where $g_i(t)$ is referred to as the in-phase component, and $g_q(t)$ is referred to as the quadrature component. Expressing 5 in this form:

$$g(t) = A_s\cos[\tfrac{1}{4}(2i - 1)\pi]\cos(2\pi f_c t) - A_s\sin[\tfrac{1}{4}(2i - 1)\pi]\sin(2\pi f_c t) \qquad (7)$$

where $g_i(t) = A_s\cos[\frac{1}{4}(2i - 1)\pi]$, and similarly $g_q(t) = A_s\sin[\frac{1}{4}(2i - 1)\pi$. Let $g_i(t)$ and $g_q(t)$ be modulated by the odd and even numbered data stream input bits respectively. Note that each quadrature component takes on values of $\pm A_s/\sqrt{2}$ depending upon the data streams.

Since the two quadrature carrier components are coherently orthogonal signals, the two binary data streams modulating the two signals can be demodulated independently. The two QPSK quadrature carrier components $g_i(t)$ and $g_q(t)$ along with their sum g(t) are shown in figure 3 for the input binary data sequence 01101000.

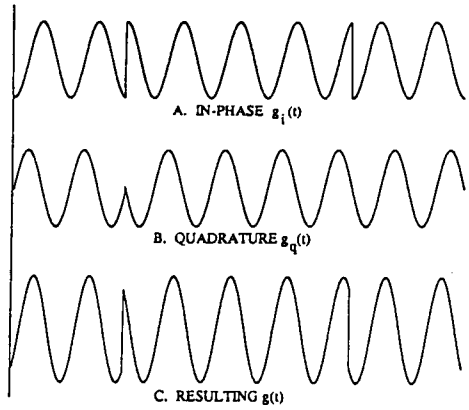


FIGURE 3
QPSK WAVEFORMS

Minimum Shift Keying (MSK)

Now consider a special case of QPSK where:

$$g(t) = A_s\cos[\theta(t)]\cos(2\pi f_c t) - A_s\sin[\theta(t)]\sin(2\pi f_c t) \qquad (8)$$

For a given bit rate $T_b = T/2$, f_c is the carrier frequency which is chosen to be an integer multiple of $1/4T_b$. $\theta(t)$ is defined as follows:

$$\theta(t) = \theta(0) \pm [\pi/(2T_b)]t, \quad 0 \le t \le T_b \qquad (9)$$

where the + corresponds to the symbol 1, the - corresponds to the symbol 0, and $\theta(0)$ is 0 or π depending upon the past history of the modulating signal. In this case $g_i(t) = \cos[\theta(t)]$, and $g_q(t) = \sin[\theta(t)]$. Let $g_i(t)$ and $g_q(t)$ be modulated by the odd and even numbered input data stream bits as before, but introduce an offset of T_b between these two data streams. Note that each quadrature component takes on maximum values of $\pm A_s/\sqrt{2}$ as before. In this case, however, the amplitudes of each quadrature component are sinusoidally weighted, and the two quadrature components never change phase at the same instant of time.

As for QPSK, the two MSK quadrature carrier components can be demodulated independently. For MSK, the two quadrature carrier components $g_i(t)$ and $g_q(t)$ along with their sum g(t) are shown in figure 4 for the input binary data sequence 01101000.

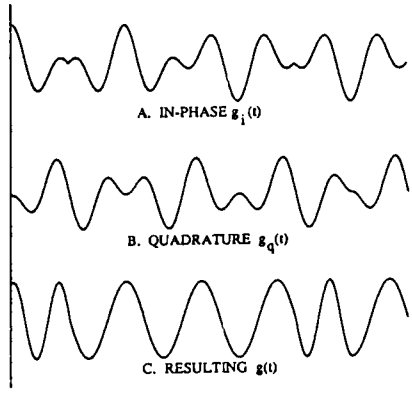


FIGURE 4
MSK WAVEFORMS

The power spectral density of any bandpass signal g(t), given by S(f) is, except for a scaling factor, a frequency shifted version of the baseband power spectral density $S_b(f)$. Therefore, the power spectral density of g(t) is given by the power spectral density of $g_i(t) + g_q(t)$. Because these two components are independent, the total power spectral density is given by the sum of the power spectral densities of $g_i(t)$ and $g_q(t)$. For a data sequence of equally probable ones and zeroes, the baseband power spectral density of g(t) in the case of QPSK is thus:

$$S_b(f) = 4E_b\left[\frac{\sin^2(2\pi T_b f)}{2\pi T_b f}\right]^2 \qquad (10)$$

Similarly, the baseband power spectral density of g(t) in the case of MSK is:

$$S_b(f) = 32E_b/\pi^2\left[\frac{\cos(2\pi T_b f)}{16T_b{}^2f^2 - 1}\right]^2 \qquad (11)$$

Equations 10 and 11 are normalized with respect to $4E_b$ and plotted as a function of $T_b f$ in figure 5.

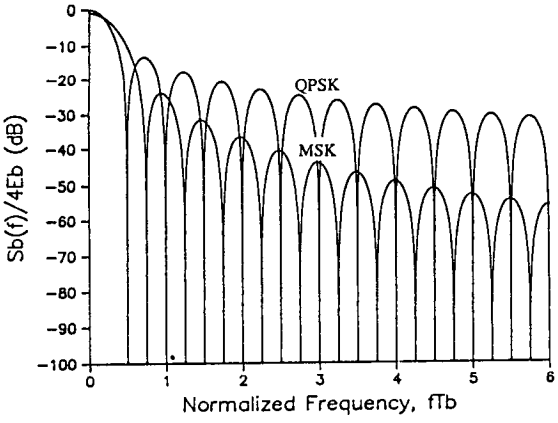


FIGURE 5
BASEBAND POWER SPECTRAL DENSITY FOR QPSK AND MSK MODULATIONS

Comparing the power spectral densities of MSK and QPSK in figure 5, it can be seen that the MSK waveform contains much less high frequency energy than the QPSK waveform. This can be explained by comparing the two waveforms in figures 3 and 4. Note that the instantaneous phase of the QPSK waveform is subject to instantaneous changes, while the instantaneous phase of the MSK waveform is continuous with time. It is the instantaneous phase changes which cause the QPSK waveform to contain more energy at the higher frequencies than the MSK waveform.

Since the removal of this high frequency energy by bandlimiting produces unwanted carrier amplitude fluctuations, it is apparent that by choosing the modulation which produces less high frequency energy, this effect can be minimized. Therefore, to minimize the deleterious effects caused to the voice modulation by the data modulation, MSK is the modulation technique of choice.

In the case of a communication channel with additive white Gaussian noise, the bit error rate of a QPSK or MSK signal can be written as: [18]

$$P(\text{error}) = 2Q[(2E_b/N_o)^{1/2}] - Q^2[(2E_b/N_o)^{1/2}] \qquad (12)$$

where $E_b = (A_s{}^2T_b)/2$ is the signal energy per bit, N_o is the noise spectral density, and Q is the Q function defined as:

$$Q(x) = (2\pi)^{-1/2}\int_x^\infty \exp(-t^2/2)\,dt \qquad (13)$$

Equation 12 is plotted as a function of E_b/N_o in figure 6.

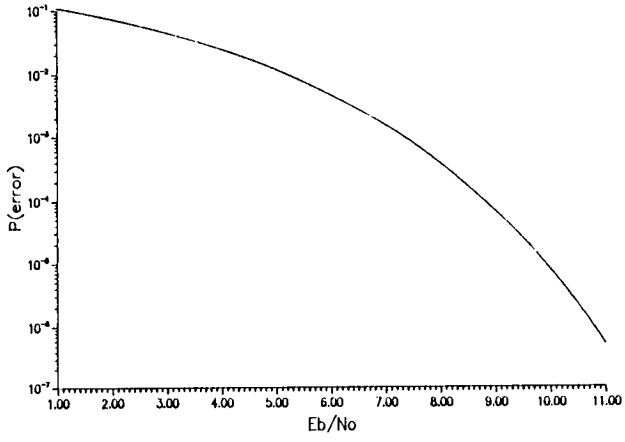


FIGURE 6
BIT ERROR PROBABILITY FOR QPSK AND MSK MODULATIONS

Analysis

A computer simulation of a combined amplitude and phase modulation communication system has been performed, and the results are available in the literature [19],[20]. The computer simulation was performed using discrete numerical techniques and sampling of the signals. Processing was performed on the equivalent baseband system. The signal was processed in blocks of 2048 samples, and was sampled at 19200 Hertz (Hz). Filtering was performed in the frequency domain, and non-linear operations such as modulation and demodulation were performed in the time domain. A Fast Fourier Transform (FFT) algorithm was used for the transformation between frequency and time domains.

The signal and system parameters were chosen to simulate transmission by hybrid modulation of voice and data on a VHF communication channel. Bandpass filtering was performed at the output of the transmitter and at the input of the receiver. The filter characteristics were as follows:

Transmitter filter - 4th order Butterworth
3 dB bandwidth of ± 7.5 kHz

Receiver filter - 8th order Butterworth
3 dB bandwidth of ± 5.0 kHz

The modulated signals of interest here were phase modulated at various data rates using MSK. The same signals were then amplitude modulated using a 4 second sample of an actual voice, and by a simulated voice signal consisting of the sum of

five tones as shown:

$$m(t) = \sum_{i=1}^{5} a_i \cos(2\pi f_i t) \qquad (14)$$

where a_i and f_i were chosen as shown in table 2. This signal and its corresponding power spectral density are shown in figures 7 and 8.

Amplitudes		Frequencies (Hz)	
a_1	2/9	f_1	468.75
a_2	1/3	f_2	937.50
a_3	2/9	f_3	1406.25
a_4	1/9	f_4	1875.00
a_5	1/9	f_5	2343.75

TABLE 2
VALUES OF a_i AND f_i FOR SIMULATED VOICE SIGNAL IN EQUATION 14

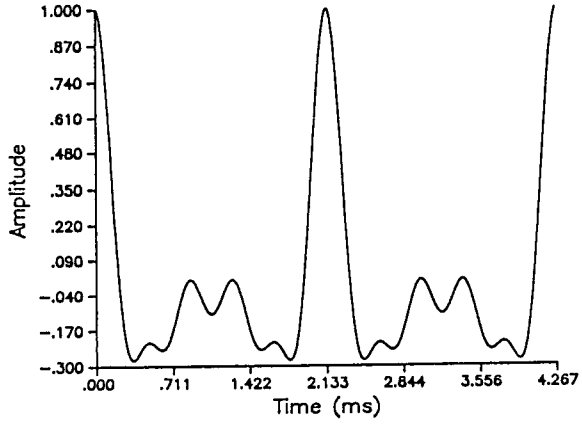


FIGURE 7
SIMULATED VOICE SIGNAL m(t)

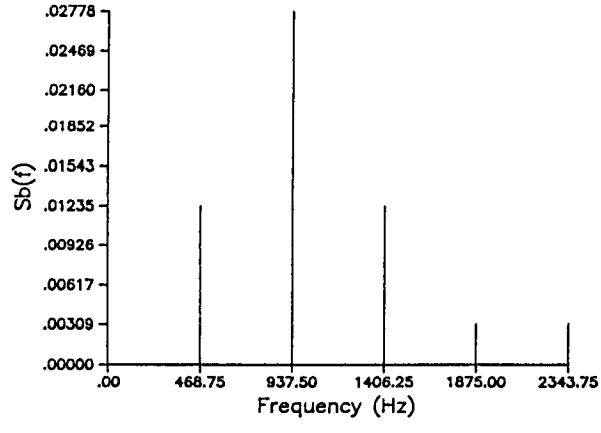


FIGURE 8
POWER SPECTRAL DENSITY OF SIMULATED VOICE SIGNAL m(t) (SAME FOR NEGATIVE FREQUENCIES)

The analysis resulted in the determination of both signal-to-noise ratio (SNR) of the detected voice signal and the bit error rate for the data signal. The voice channel is assumed noiseless such that the noise in the channel is due entirely to the effects of the data modulation. For the data signal, additive white Gaussian noise is introduced in the channel, and the overall bit error rate including the effects of the amplitude modulation is experimentally determined.

This analysis does not take into account amplitude limiting of the phase modulated portion of the carrier. Such limiting would occur in some Class C transmitter power amplifiers, and would be necessary to reduce the dynamic range of the signal presented to the phase detector in the receiver. Instead, the effects of phase detecting the amplitude modulated version of the phase modulated carrier are considered.

Analysis Results

The voice signal SNR results for the simulated voice with μ set to 0.8 are shown in figure 9 as a function of data rate. The same data for the actual voice signal with μ set to 0.9 are shown figure 10. The results for the data modulation show an overall degradation of 2-4 dB in E_b/N_o as compared to an ideal infinite bandwidth and constant amplitude data modulation.

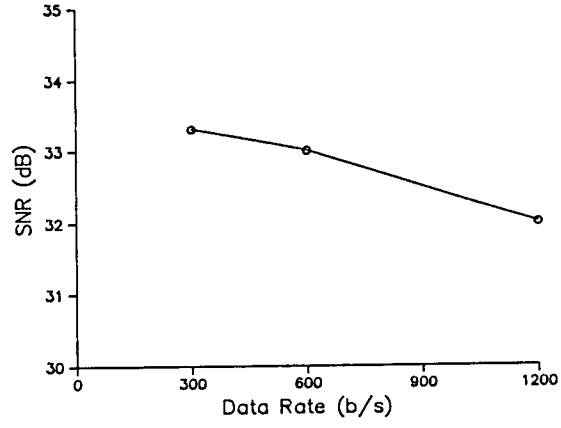


FIGURE 9
SNR RESULTS FOR HYBRID MODULATION USING SIMULATED VOICE SIGNAL (u = 0.8)

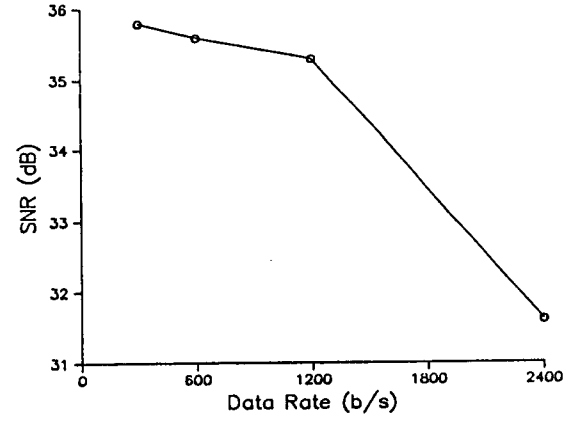


FIGURE 10
SNR RESULTS FOR HYBRID MODULATION USING ACTUAL VOICE SIGNAL (u = 0.9)

Further Analysis

If both the amplitude and bandwidth limiting effects on the data modulation are considered, further analysis can be applied to predict the

overall reduction in bit error rate. Such analysis has been performed for various cases of filter bandwidths assuming an ideal limiter characteristic [21]. The results show that despite severe distortion of data symbol shapes and the introduction of interphasor crosstalk, E_b/N_o is degraded by a maximum of 0.5 dB.

With amplitude limiting, an additional degradation of E_b/N_o can be expected since the amplitude of the data modulated carrier would be limited to its lowest value. For a carrier with constant amplitude A_c, the energy per bit is given by:

$$E_b = A_c^2 T_b/2 \qquad (15)$$

However, if the carrier amplitude is modulated by a voice signal with some known value of μ, equations 3 and 4 show that the minimum value of the overall bit energy would be given by:

$$E_{bmin} = A_c^2(1 - \mu)^2 T_b/2 \qquad (16)$$

Since the limiter would provide this level to the phase detector at all times, a reduction in E_b/N_o, in decibels, would be given by:

$$P_{loss} = 10\log(E_{bmin}/E_b) = 20\log(1 - \mu) \qquad (17)$$

For μ of 0.8, this would represent a loss of approximately 14 dB. Adding this amount to the worst-case degradation predicted due to the presence of amplitude modulation and limiting of both amplitude and bandwidth gives a total loss in E_b/N_o of:

$$14 + 4.0 + 0.5 = 18.5 \text{ dB}$$

relative to a constant amplitude carrier. This loss would result in a decreased distance of communication for acceptable system performance.

It is important to consider the spectrum requirements of the overall hybrid modulated waveform. Because the amplitude and phase modulations are statistically independent, the overall power spectral density at baseband can be found by the convolution of the power spectral densities of the voice modulation process and the phase modulation process. The baseband power spectral density of the MSK modulated carrier is given by equation 11 and shown in figure 5. If the simulated voice signal given by equation 14 is chosen to represent the voice information, then the baseband power spectral density of the voice modulated carrier is as shown in figure 8. The convolution of the two signals at the carrier frequency yields the power spectral density shown in figure 11. Also shown on the plot are the Federal Communications Commission (FCC) spectrum limits for the VHF aeronautical communication band [22]. The hybrid modulated waveform conforms to this requirement.

Data Redundancy Reduction

The transmission time required for the uplink of a complete weather data graphical image at a fixed data rate can be substantially reduced using image redundancy reduction techniques. Many methods exist for the accomplishment of this redundancy reduction.

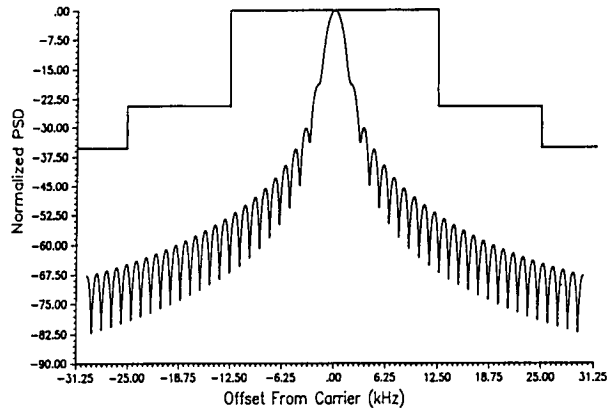


FIGURE 11
NORMALIZED POWER SPECTRAL DENSITY OF HYBRID MODULATED WAVEFORM USING SIMULATED VOICE SIGNAL (SHOWN WITH VHF VOICE CHANNEL SPECTRUM LIMITS)

The method chosen for data redundancy reduction of graphical weather images will depend upon the characteristics of the image. For the uplink of weather radar reflectivity patterns, however, run-length encoding provides very high compression ratios and is implemented more easily than many types of image compression. Run length encoding codes runs of repeated pixel values with a starting location and a count of the number of repeated pixels. Because NWS weather radar reflectivity patterns contain only six intensity levels and these levels are usually grouped together to form weather cells, repeated pixel runs occur frequently. Thus run length encoding very efficiently compresses the amount of data necessary for the description of these images. Further compression can be achieved by not transmitting weather map overlays. Since this information is constant, and digital data storage costs are decreasing, the weather map information can be stored in the cockpit and overlaid on the current weather data.

Radar reflectivity patterns which include weather activity over large areas of the radar coverage have been compressed to 2000 bytes using run length coding and suppression of map detail [23]. Transmitting this data at 2400 bits/s would result in a total transmission time of less than 7 seconds.

Conclusions

Some requirements for the uplink of weather data to the aircraft have been delineated, and it has been shown that the Mode S Beacon System data link capability will not be sufficient to provide weather data uplink capability to all airspace users.

A system capable of meeting weather data uplink requirements has been described. The system uses hybrid modulation techniques for the transmission of both voice and data on an existing aeronautical VHF voice communication channel. The system is particularly well suited for the provision of weather data products to the airspace user in the terminal area where Mode S capabilities have been shown to be inadequate.

A computer analysis shows that the hybrid modulation technique would result in an E_b/N_o loss of up to 18.5 dB as compared to a phase modulated constant amplitude carrier. This loss in E_b/N_o results in an increased bit error probability as shown in figure 6. By limiting communication distance, however, acceptable performance can be maintained. The voice signal is shown to provide an SNR of approximately 30 dB when combined with the MSK phase modulation. The MOPS for VHF airborne communication equipment establishes 25 dB as the required minimum SNR when the received signal strength is well above the receiver noise floor [24]. Based on this requirement, the SNR obtained using the hybrid modulation technique would be adequate. The hybrid modulation technique would thus provide acceptable performance if the AM modulation factor is limited to the range of 0.7 to 0.9,and MSK is used for data transmission at a data rate of 2400 bits/s.

References

1. National Transportation Safety Board, "Annual Review of Aircraft Accident Data, U.S. General Aviation, Calendar Year 1985", NTSB/ARG-87/03, Washington, D.C., October 13, 1987, p.34.

2. National Transportation Safety Board, "Annual Review of Aircraft Accident Data, U.S. Air Carrier Operations, Calendar Year 1985", NTSB/ARC-87/03, Washington, D.C., November 27, 1987, p.23.

3. National Transportation Safety Board, "Aircraft Accident Report, Delta Airlines, Inc., Lockheed L-1011-385-1, N726DA, Dallas/Fort Worth International Airport, Texas, August 2, 1985", NTSB/AAR-86/05, Washington, D.C., August 15, 1986, p.84.

4. U.S. Department of Transportation, Federal Aviation Administration, "Aviation Weather System Plan", Washington, D.C., 1984, p.5-2.

5. Ibid., pp.3-1-12.

6. Parker, C.B., "Aircraft Weather Data Dissemination System", Proceedings of the Joint University Program for Air Transportation Research-1987, National Aeronautics and Space Administration, NASA CP-3028.

7. McFarland, R.H., "A Delineation of Critical Weather Factors Concerning General Aviation", OU/AEC/EER 53-3, Ohio University, Avionics Engineering Center, Athens, Ohio, November 1981.

8. McFarland, R.H., "An Experimental Investigation of the Efficacy of Automated Weather Data Transmission to Aircraft in Flight", DOT/FAA/PM-83/11, U.S. Department of Transportation, Federal Aviation Administration, Program Engineering and Management Service, Washington, D.C., December 1982.

9. Hansman, R.J., Chamon, C., Peltzman, S., and Lakhani, H., "Evaluation of Advanced Microburst Cockpit Displays", presented at FAA/NASA Joint University Program for Air Transportation Research Conference, Cambridge, Massachusetts, September 1987.

10. Orlando, V.A. and Drouilhet, P.R., "Mode S Beacon System: Functional Description", DOT/FAA/RD-82/52, Department of Transportation, Federal Aviation Administration, Systems Research and Development Service, Washington, D.C., October 27, 1982.

11. Radio Technical Commission for Aeronautics, "Minimum Operational Performance Standards For Air Traffic Control Radar Beacon System/Mode Select (ATCRBS/Mode S) Airborne Equipment", RTCA/DO-181, Washington, D.C., March 25, 1983.

12. Op.cit., DOT/FAA, "Aviation Weather System Plan", pp.5-15-17.

13. Op.cit., RTCA, "Minimum Operational Performance Standards for ATCRBS/Mode S", p.42.

14. U.S. Department of Transportation, Federal Aviation Administration, "National Airspace System Plan, Facilities, Equipment and Associated Development", Washington, D.C., April 1984, pp.IV-78-9.

15. Federal Communications Commission, Rules and Regulations, Part 87 - Aviation Services, Washington, D.C., September 1987, p.15.

16. Haykin, S., Communication Systems, 2nd ed., John Wiley and Sons, New York, 1983.

17. Pasupathy, S., "Minimum Shift Keying: A Spectrally Efficient Modulation", IEEE Communications Society Magazine, vol.17, no.11, July 1979, pp.14-22.

18. Op.cit., Haykin, pp.555-558.

19. Benelli, G. and Fantacci, R., "An Integrated Voice-Data Communication System for VHF Links", IEEE Transactions on Communications, vol.COM-31, no.12, December 1983, pp.1304-8.

20. Benelli, G., "VHF Radio Link for Ground-Air-Ground Communications Using an Integrated Voice-Data Modulation", Electronics Letters, vol.18, no.13, June 24, 1982, pp.555-6.

21. Mathwich, H.R., Balcewicz, J.F., Hecht, M., "The Effect of Tandem Band and Amplitude Limiting on the E_b/N_o Performance of Minimum (Frequency) Shift Keying (MSK)", IEEE Trans. on Comm., vol.COM-22, no.10, October 1974, pp.1525-40.

22. Op.cit. Federal Communications Commission, Rules and Regulations, Part 87, p.14.

23. Kelley, J.P., "A System Concept for the Collection and Dissemination of Weather Radar Data Using Satellite Communications", MTR-87W15, The MITRE Corporation, McLean, Virginia, January 1987, p.3-4.

24. Radio Technical Commission for Aeronautics, "Minimum Operational Performance Standards for Airborne Radio Communications Equipment Operating Within the Radio Frequency Range 117.975-137.000 MHz", RTCA/DO-186, Washington, D.C., January 1984, p.10.

This work was supported by the Federal Aviation Administration and the National Aeronautics and Space Administration under grant NGR 36-009-017.

PRINCETON UNIVERSITY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT PRINCETON UNIVERSITY, 1988-1989

Robert F. Stengel
Department of Mechanical and Aerospace Engineering
Princeton University
Princeton, New Jersey

SUMMARY OF RESEARCH

The Air Transportation Technology Program at Princeton University, a program emphasizing graduate and undergraduate student research, proceeded along seven avenues during the past year:

- Microburst and Wind Shear Hazards to Aircraft
- Intelligent Flight Control Systems
- Real-time Cockpit Simulation
- Revitalization of General Aviation
- Alternatives for the Air Traffic Control System
- Stochastic Robustness of Flight Control Systems
- Neural Networks for Flight Control

It has become apparent that severe downdrafts and resulting high velocity outflows caused by microbursts present a significant hazard to aircraft on takeoff and final approach. Microbursts, which are often associated with thunderstorm activity, also can occur in the vicinity of dissipating convective clouds that produce no rainfall at ground level. Microburst encounter is a rare but extremely dangerous phenomenon that accounts for one or two air carrier accidents and numerous general aviation accidents each year (on the average). Conditions are such that an aircraft's performance envelope may be inadequate for safe penetration unless optimal control strategies are known and applied.

While a number of simulation studies have been directed at the problem, there are varied opinions in the flying community regarding the best piloting procedures, and optimal control strategies have only recently been defined. Graduate student Mark Psiaki completed a study of optimal trajectories for penetration of microbursts when encounter is unavoidable. His initial work (reported earlier) showed that simple control laws could greatly reduce an aircraft's response to wind shear. His 1987 Ph.D. thesis

presents flight envelopes within which optimal application of throttle and pitch control produce acceptable deviations from nominal takeoff and landing flight paths for a jet transport and a propeller-driven general aviation aircraft. Although the response mechanism is the same, jet transport and general aviation aircraft behave somewhat differently in microbursts; the larger, heavier aircraft are more adversely affected by variations in the horizontal wind, while the smaller, lighter aircraft have greater difficulty with the downdraft. These findings are summarized in Ref. 1.

The emphasis of current wind shear research, being pursued by graduate student Alex Stratton under a separate NASA grant, is on developing an expert system for wind shear avoidance [2,3]. Our principal objectives are to develop methods for assessing the likelihood of wind shear encounter (based on real-time information in the cockpit), for deciding what flight path to pursue (e.g., abort, go-around, normal climbout, or glide slope), and for using the aircraft's full potential to combat wind shear. This study requires the definition of deterministic and statistical techniques for fusing internal and external information, for making "go/no-go" decisions, and for generating commands to the aircraft's autopilot and flight directors in automatic and manually controlled flight.

Undetected system failures and/or inadequately defined recovery procedures have contributed to numerous air carrier incidents and accidents. The infamous DC-10 accident at Chicago's O'Hare Airport, in which loss of an engine pod, subsequent loss of subsystems, and asymmetric wing stall led to disaster, provides a prototype for the kind of tragedy that could be averted by intelligent flight control systems. (An intelligent control system is one that uses artificial intelligence concepts, e.g., an expert systems program, to improve performance and failure tolerance.) Although many methods of modern control theory are applicable, the scope of the problem is such that none of the existing theories provides a complete and practical solution to the problem. At the same time, heuristic logic may be applicable, but it has yet to be stated in satisfactory format.

Graduate student David Handelman has developed a knowledge-based reconfigurable flight control system that is implemented with the Pascal programming language using parallel microprocessors. This expert system could be considered a prototype for a failure-tolerant control system that can be constructed using existing hardware. The knowledge-based flight control system is specified initially and tested using the LISP programming language. When desired logic is determined, the corresponding Pascal code is generated automatically. Concepts for rule-based learning and

control that can be used in such systems are contained in two technical papers and in Mr. Handelman's Ph.D. thesis [4 - 6].

Helping a pilot make quick decisions under high workload conditions is important for aircraft missions of all types. In research principally supported by an Army/Navy grant but reported at numerous quarterly reviews of the Joint University Program, Brenda Belkin has developed an expert system of expert systems called AUTOCREW. In her M.S.E. thesis [7], Ms. Belkin uses the paradigm of a hypothetical aircraft crew to facilitate the assignment of tasks, rules, and data within parallel knowledge bases. AUTOCREW performs a cyclical search in which the Director expert system, the electronic analog of the aircraft commander, establishes goals that invoke the crew-member expert systems. The crew members then perform such tasks as observation, monitoring, and control in response to continuing needs as well as special requests from the Director. The application of AUTOCREW to civil jet transports is apparent, and Ms. Belkin has participated in the Joint University Program with this in mind.

Development of a real-time cockpit simulator continued during the year. The simulator provides a single-person crew station with both conventional and advanced control devices; it currently is programmed to simulate the Navion single-engine general aviation airplane. Displays for the "out-the-window" view and the control panel devices are generated entirely by computers and are presented on color cathode-ray tubes. The external view is generated and displayed by a Silicon Graphics IRIS 3020 Workstation, which has 1024 x 768-pixel resolution and can perform over 85,000 3-D transformations per second. A central computing unit performs dynamic and control calculations, accepts analog inputs, drives the panel displays, and commands the external view. It is a special-purpose Multibus computer employing parallel 80286 processors and special-purpose graphics boards, which are controlled by an IBM PC-AT computer. Senior Jeffrey Seinwill developed an all-attitude artificial horizon display for the simulator during the year [8].

For all practical purposes, the production of small general aviation (GA) airplanes in the U.S. has come to an end. Fewer than a thousand four-seaters were produced in 1987, and most of the principal American GA manufacturers have all but abandoned the business. Although there are many reasons for the erosion of this industry, outmoded technology is a major factor. General aviation has a potentially significant role to play not only in the National Transportation System but in the economic infrastructure of the country. Revival demands reinvention of the small plane, in

what might be called a Modern-Equipment General Aviation airplane. There is a wealth of fundamental research to be done in this area, much of it already under way in the Air Transportation Technology Program. Senior Joakim Karlsson's independent work project dealt with the design of such an aircraft [9] using modern aerodynamic, structural, and control technology. Senior Bruce Weiner conducted an economic review of the price-performance characteristics of small aircraft during the period from 1970 to 1987 that graphically illustrates the effects of increased product liability costs [10].

Air commerce is a significant factor in the nation's economy, involving hundreds of millions of passenger trips and tens of billions of dollars in operating revenue each year. Recent revenue growth rates average 10 percent per year; much of this growth can be traced to the deregulation of the nation's airlines in 1978. This growth has caused severe air traffic control problems, most often demonstrated in flight delays and reflected by growing concerns for passenger safety. Despite the aggregate economic benefits of deregulation, many people are questioning whether the benefits outweigh the perceived costs of reduced confidence and increased delays. Senior Lawrence Neubauer analyzes alternatives for the air traffic control system in his thesis, giving particular attention to the institutional structure of the FAA [11].

Control system robustness is defined as the ability to maintain satisfactory stability or performance characteristics in the presence of all conceivable system parameter variations. While assured robustness may be viewed as an alternative to gain adaptation or scheduling to accommodate known parameter variations, more often it is seen as protection against uncertainties in plant specification. Consequently, a statistical description of control system robustness is consistent with what may be known about the structure and parameters of the plant's dynamic model. Guaranteeing robustness has long been a design objective of control system analysis, although in most instances, insensitivity to parameter variations has been treated as a deterministic problem. Graduate student Laura Ryan is investigating a simple numerical procedure for estimating the *stochastic robustness* of control systems [12-14]. Monte Carlo evaluation of the system's eigenvalues allows the *probability of instability* and the related *stochastic root locus* to be estimated. This analysis approach treats not only Gaussian parameter uncertainties but non-Gaussian cases, including uncertain-but-bounded variations. Trivial extensions of the procedure admit alternate discriminants to be considered. Thus, the probabilities that stipulated degrees of instability will be exceeded or that closed-loop roots will leave

desirable regions also can be estimated. Results are particularly amenable to graphical presentation.

There is growing interest in the use of neural networks for computational decision-making and control, brought about by the advent of small, fast, inexpensive computers. The neural network paradigm offers a potentially attractive structure for flight control systems that adapt to changing flight conditions and system failures, but much is to be learned about the practicality of such an approach. Graduate student Dennis Linse has begun to examine this potential. To date, he has formulated various multi-layer networks to trim the longitudinal motions of a twin-jet transport. The network is resilient in the face of elemental failures and provides moderately accurate open-loop control of throttle and elevator over a wide range of flight conditions.

ANNOTATED REFERENCES OF 1988-89 PUBLICATIONS

1. Psiaki, M., and Stengel, R.F., "Performance Limits for Optimal Microburst Encounter," *Proceedings of the AIAA 15th Atmospheric Flight Mechanics Conference*, Minneapolis, Aug 1988, pp. 358-370.

The performance limits of optimal aircraft control strategies for microburst encounter are presented. The purpose is to determine the "edges of the envelope" for no-accident aircraft penetration of microburst wind shear. Over 1,100 optimal trajectories have been computed for jet transport and general aviation aircraft flying through idealized microbursts. They have been generated using a Successive Quadratic Programs trajectory optimization algorithm, which directly handles inequality constraints. Qualitative aspects of the best strategies provide a composite picture of good control in a microburst. Variations of the no-accident performance with microburst type, intensity, length scale, and location define performance limits.

Optimal performance limits show three length-scale regimes. At short length scales, hazards usually associated with gustiness predominate. At intermediate length scales, a degraded ability to maintain flight path and/or vertical velocity sets the limiting microburst intensities for no-accident performance. At very long microburst length scales, the hazards associated with intense steady winds are the critical safety limits. The abil-

ity to successfully transit a microburst also varies strongly with microburst location. The performance limits show that both aircraft types, if controlled properly, can penetrate very severe microbursts. Nevertheless, even the best control strategies have their limits. The jet transport performance limits occur at higher microburst intensities than the general aviation limits.

2. Stengel, R.F., and Stratton, D.A., "An Expert System for Wind Shear Avoidance," presented at the Second Combined Manufacturers and Technology Airborne Wind Shear Review Meeting, Williamsburg, VA, Oct 1988.

Flight in strong wind shears, especially microbursts, poses a unique and severe hazard to aircraft. The disturbance caused by the wind field may literally exceed the performance characteristics of the aircraft, making safe transit impossible even with optimal guidance and control strategies. An unusual degree of piloting skill may be required to successfully elude danger. Nevertheless, planes fly in moderate wind shear all the time; pilots learn to handle crosswinds, gustiness, and moderate frontal activity. The problem is that microbursts are random, rare phenomena; pilots do not develop the needed skills for coping with wind shear through normal experience. The typical pilot is likely to be confronted with a life-threatening wind shear only once or twice during an entire flying career; hence, it is unlikely that he or she can learn all the important signs of wind shear and maintain a high level of proficiency in the proper control procedures.

On-board computation provides an excellent opportunity to assist the pilot in surviving encounters with severe wind shears, but the logic that must be executed in real time is complex and must have sufficient inputs for framing decisions about appropriate control actions. The computer program(s) and hardware to perform this task must have attributes of expert systems and control systems, they must account for the limitations of aircraft performance, and they must operate in real time. At least as important as its technical specifications, the on-board system must provide a satisfactory interface with the flight crew, which bears the ultimate responsibility for assuring safety. This means not only that the system must deduce near-optimal strategies and tactics for emergency situations but that it must distinguish between truly hazardous conditions and the more likely alternatives associated with normal aircraft operations.

A program to investigate ways of protecting against wind shear has begun at Princeton University, with current emphasis on developing an

expert system for wind shear avoidance. Our principal objectives are to develop methods for assessing the likelihood of wind shear encounter (based on real-time information in the cockpit), for deciding what flight path to pursue (e.g., abort, go-around, normal climbout, or glide slope), and for using the aircraft's full potential to combat wind shear. This study requires the definition of deterministic and statistical techniques for fusing internal and external information, for making "go/no-go" decisions, and for generating commands to the aircraft's autopilot and flight directors in automatic and manually controlled flight.

The expert system for pilot aiding is based on results of the Integrated FAA Wind Shear Program. Its two-volume manual presents an overview, pilot guide, training program, and substantiating data and provides guidelines for this initial development. The WindShear Safety Advisor currently being developed at Princeton contains over 140 rules and is coded in the LISP programming language for implementation on a Symbolics 3670 LISP Machine.

3. Stengel, R.F., and Stratton, D.A., "An Expert System for Wind Shear Avoidance," presented at the 1989 American Control Conference, Pittsburgh, June 1989.

A study of intelligent guidance and control concepts for protecting against the adverse effects of wind shear during aircraft takeoffs and landings is being conducted, with current emphasis on developing an expert system for wind shear avoidance. Principal objectives are to develop methods for assessing the likelihood of wind shear encounter (based on real-time information in the cockpit), for deciding what flight path to pursue (e.g., takeoff abort, landing go-around, or normal climbout or glide slope), and for using the aircraft's full potential for combating wind shear. This study requires the definition of both deterministic and statistical techniques for fusing internal and external information, for making "go/no-go"decisions, and for generating commands to the aircraft's autopilot and flight directors for both automatic and manually controlled flight. The program has begun with the development of the WindShear Safety Advisor, an expert system for pilot aiding that is based on the FAA Windshear Training Aid, a two-volume manual that presents an overview, pilot guide, training program, and substantiating data provides guidelines for this initial development. The WindShear Safety Advisor expert system currently contains over 200 rules and is coded in the LISP programming language.

4. Handelman, D.A., and Stengel, R. F., "Rule-Based Mechanisms of Learning for Intelligent Adaptive Flight Control," *Proceedings of the 1988 American Control Conference*, Atlanta, June 1988, pp. 208-213.

This paper investigates how certain aspects of human learning can be used to characterize learning in intelligent adaptive control systems. Reflexive and declarative memory and learning are described. It is shown that model-based systems-theoretic adaptive control methods exhibit attributes of reflexive learning, whereas the problem-solving capabilities of knowledge-based systems of artificial intelligence are naturally suited for implementing declarative learning. Issues related to learning in knowledge-based control systems are addressed, with particular attention given to rule-based systems. A mechanism for real-time rule-based knowledge acquisition is suggested, and utilization of this mechanism within the context of failure analysis for fault-tolerant flight control is demonstrated.

5. Handelman, D.A., and Stengel, R. F., "Perspectives on the Use of Rule-Based Control," *Proceedings of the IFAC Workshop on Artificial Intelligence in Real-Time Control*, Swansea, UK, Sept 1988.

This paper addresses issues regarding the application of artificial intelligence techniques to real-time control. Advantages associated with knowledge-based programming are discussed. A proposed rule-based control technique is summarized and applied to the problem of automated emergency procedure execution. Although emergency procedures are by definition predominantly procedural, their numerous evaluation and decision points make a declarative representation of the knowledge they encode highly attractive, resulting in an organized and easily maintained software hierarchy. Simulation results demonstrate that real-time performance can be obtained using a microprocessor-based controller. It is concluded that a rule-based control system design approach may prove more useful than conventional methods under certain circumstances, and that declarative rules with embedded procedural code provide a sound basis for the construction of complex control systems.

6. Handelman, D.A., "A Rule-Based Paradigm for Intelligent Adaptive Flight Control," Ph. D. Thesis, Princeton University, Report No. 1858-T, April 1989.

Although fault tolerance always has been an important aspect of aircraft design, the reduced static stability and increased maneuverability of modern aircraft complicate the problem by shortening the amount of time available to detect, identify, and adjust for component failures. This dissertation investigates the use of highly integrated symbolic and numeric processing in real-time knowledge-based systems for enhanced automatic aircraft failure accommodation. A rule-based control technique is proposed whereby procedural activity is attained through the manipulation of declarative expressions. Rules are used to encode common-sense dependencies, to incorporate expert knowledge on specific situations, and to invoke algorithmic mathematical procedures. Task execution occurs as a by-product of search through these knowledge-base rules. Also proposed is a rule-based controller development system that utilizes a high-level symbolic LISP environment for preliminary system design. Automatic LISP-to-Pascal knowledge-base translation is used to provide dramatically increased execution speed and an environment for highly integrated symbolic and numeric computation. The utility of the control technique is demonstrated through the construction of a multi-processor Rule-Based Flight Control System (RBFCS) for a CH-47 tandem-rotor helicopter. The RBFCS is shown to accommodate multiple simulated failures affecting the electrical, hydraulic, and stability augmentation subsystems of the helicopter. It is concluded that declarative rules with embedded procedural code provide a sound basis for the construction of complex control systems.

7. Belkin, B.L., "Cooperative Rule-Based Systems for Aircraft Navigation and Control," M.S.E. Thesis, Princeton University, Report No. MAE-1856-T, June 1989.

This thesis proposes a design methodology for the development of multiple cooperating rule-based systems for aircraft. Nine systems, collectively called AUTOCREW were designed to automate functions and decisions. The organization of tasks is described, details of knowledge-base development and implementation are given, and performance metrics for evaluating the workload of each knowledge base are demonstrated. Several test scenarios were evaluated using an interactive graphical simulation on an IBM PC-AT computer. Software tools developed to aid in high-level design also are described. Results show that these tools facilitate rapid

prototyping of a complex system exhibiting knowledge-base cooperation, satisfactory logic flow, and human pilot-AUTOCREW interaction.

Design of one of the component expert systems, the Navigation Sensor Management (NSM) module, was pursued in considerable detail. This problem was chosen because it presented the challenge of designing an expert system from simulation data; that is, from quantitative test results. The NSM Expert was systematically derived from Kalman filter covariance data for simulated missions flown with seven different navigation system types. This development used Analysis of Variance (ANOVA) and the "ID3" algorithm. The function of the NSM Expert was to determine optimal navigation strategies from a set of available sensors based on a root-sum-square metric. Results show that the NSM Expert predicted position accuracy between 65 and 100 percent of the time for a specified "navaid" configuration and aircraft trajectory; hence, this decision-making logic could be incorporated in a scheme for best navaid selection. The systematic nature of the ANOVA/ID3 method makes it broadly applicable to expert system design when experimental or simulation data are available.

8. Seinwill, J., "Design and Testing of Cockpit Displays for a Flight Simulator," Princeton University Senior Independent Work Project Report, May 1988.

A program to simulate an artificial horizon display on a CRT was developed and tested. The code exhibits real-time performance using a Matrox Multibus graphics board to generate an image of the aircraft's pitch and roll attitudes in maneuvering flight. The display can be masked to yield circular or rectangular format, and additional alphanumeric or graphical information can be presented within the mask. A pitch "ladder" can be superimposed on the artificial horizon to show quantitative data.

9. Karlsson, J., "Preliminary Design for a Modern Equipment General Aviation (M.E.G.A.) Aircraft," Princeton University Senior Independent Work Project Report, May 1988.

A single-engine/propeller aircraft design is proposed to revitalize the U.S. general aviation industry. The objective of the design is to improve safety over existing designs, improve cost efficiency, and to provide easy-to-fly aircraft, while maintaining good performance qualities. The design approach is defined, followed by the resulting design procedure, and a preliminary design is presented. These objectives are realized through the

use of modern, but proven, technology, as well as a high degree of system redundancy.

The Modern Equipment General Aviation (M.E.G.A.) airplane is designed to carry a payload of one pilot and three passengers, as well as two children on a folding seat or 100 kg of baggage. The range specification is 1000 km, the cruising speed is 90 m/s, and the landing speed is 25 m/s. The resulting design incorporates a stratified-charge rotary engine and three lifting surfaces: a laminar-flow wing, a T-tail, and a small flapped canard in the propeller slipstream. All lifting surfaces have mechanically uncoupled controls. The wing has 50%-span simple flaps, and the wing tips are sheared to improve aerodynamic efficiency. Large cockpit windows afford good visibility, and limited use is made of composite materials to reduce weight and insure partial laminar flow over the wing.

The first phases of the preliminary design were based on conventional analytical procedures. Two computer programs, *LinAir* and *Paper Airplane*, were used to optimize design details such as wing twist and static stability to produce the final design.

10. Weiner, B.W., "Public Policy and Its Impact on The Decline Of General Aviation: Directions For Future Improvement," Senior Thesis, Woodrow Wilson School of Public and International Affairs, Princeton University, May 1988.

General aviation has not participated in the recent upswings in the economy. This year represented the ninth consecutive year of the decline in shipments of U.S.-built general aviation airplanes. The number of private pilots has been following a downtrend as well. The cause of the decline and future directions for general aviation are the topics of this thesis.

The purpose of the inquiry is to examine some of the external influences driving general aviation pilots and aircraft out of American skies. It begins by examining the benefits of general aviation in both quantitative and qualitative terms. It considers the effect that the FAA has had and the policy changes that would benefit general aviation. It develops a "hedonic" pricing model to examine the economic demand for general aviation aircraft, focusing on single-reciprocating-engine configurations. The thesis concludes with policy recommendations.

The pricing model is a function of cruise speed, range, useful load, rate of climb, landing gear (retractable or not), and aircraft age. An eight-term polynomial was found to fit the logarithm of current average price with an R^2 statistic of 0.93 for a 17-year period beginning in 1970. The resulting index was applied to the general aviation fleet on an averaged, annualized basis to determine the relative quality of available aircraft. This measure suggests that both quality and aircraft price increased during the period, although there was a downturn in both during the final two years. The price index tracked the gross national product rather well until the same point in time. The author presents evidence that this relative reversal is due to skyrocketing product liability costs, which impact established manufacturers with many aircraft in the field most heavily. Liability costs have smaller impact on foreign competitors, who do not have as large an American fleet and thus have fewer claims. Tort reform is seen as a principal means of solving this problem.

11. Neubauer, L. A., "The Air Traffic Control System: Problems and Prospects," Senior Thesis, Woodrow Wilson School of Public and International Affairs, Princeton University, April 1988.

Air commerce is a significant factor in the nation's economy, involving hundreds of millions of passenger trips and tens of billions of dollars in operating revenue each year. Recent revenue growth rates average 10 percent per year; much of this growth can be traced to the deregulation of the nation's airlines in 1978. This growth has caused severe air traffic control problems, most often demonstrated in flight delays and reflected by growing concerns for passenger safety. Despite the aggregate economic benefits of deregulation, many people are questioning whether the benefits outweigh the perceived costs of reduced safety and increased delays. Elected officials have responded by holding numerous hearings and introducing bills providing for a range of alternatives from total re-regulation to dismantling the Federal Aviation Administration.

The FAA faces a growing crisis of confidence in its ability to provide air traffic control. In response to funding, procurement, staffing, and oversight problems, three alternative solutions were analyzed in this thesis: reestablishing the FAA as an independent agency, setting up a government-owned corporation to run air traffic control, and privatizing air traffic control. After extensive review, it was concluded that the first alternative provides the best opportunity for resolving this dilemma, while creating the fewest new problems. An independent FAA would provide the opportunity for a revitalization of the Agency by reducing funding concerns,

cutting outside influence over the organization, and providing the Administrator with more flexibility to appoint staff and to recognize meritorious service. Resolution of the concerns facing air traffic control depends upon the joint efforts of FAA, DOT, Congress, and the Administration, as well as the users of the system.

12. Stengel, R.F., and Ryan, L.E., "Stochastic Robustness of Linear Control Systems," presented at the 1989 Conference on Information Sciences and Systems, The Johns Hopkins University, Baltimore, Mar 1989.

A simple numerical procedure for estimating the *stochastic robustness* of a linear, time-invariant system is described. Monte Carlo evaluation of the system's eigenvalues allows the *probability of instability* and the related *stochastic root locus* to be estimated. This definition of robustness is an alternative to existing deterministic definitions that address both structured and unstructured parameter variations directly. This analysis approach treats not only Gaussian parameter uncertainties but non-Gaussian cases, including uncertain-but-bounded variations. Trivial extensions of the procedure admit alternate discriminants to be considered. Thus, the probabilities that stipulated degrees of instability will be exceeded or that closed-loop roots will leave desirable regions also can be estimated. Results are particularly amenable to graphical presentation.

13. Stengel, R.F., and Ryan, L.E., "Stochastic Robustness," presented at the SIAM Conference on Control in the 90's," San Francisco, May 1989.

This paper is an oral-only extension of the above paper with extended numerical results for the Doyle LQG counterexample. In particular, the comparison of stochastic robustness with singular-value analysis is pursued; whereas the latter provides only qualitative guidelines for "loop transfer recovery," the former provides a specific solution for the amount of robustness recovery required as a function of parameter uncertainty.

14. Stengel, R.F., and Ryan, L.E., "Multivariate Histograms for Analysis of Linear Control System Robustness," presented at the 1989 American Control Conference, Pittsburgh, June 1989.

A simple numerical procedure for estimating the *stochastic robustness* of a linear, time-invariant system is described. Based on Monte Carlo evaluation of the system's eigenvalues, this analysis approach introduces the

probability of instability as a scalar measure of stability robustness. The related *stochastic root locus*, a portrayal of the root probability density, provides insight into robustness characteristics. Parameter variations are not limited to Gaussian parameter uncertainties; non-Gaussian cases, including uncertain-but-bounded variations, can be considered as well. Confidence intervals for the scalar probability of instability address computational issues inherent in Monte Carlo simulation. An example demonstrates stochastic robustness as applied to an aircraft control system in which parameters are alternately considered to have Gaussian, uniform, or binary probability distributions.

APPLICATION OF STOCHASTIC ROBUSTNESS TO AIRCRAFT CONTROL SYSTEMS

Laura E. Ryan
Department of Mechanical and Aerospace Engineering
Princeton University

INTRODUCTION

Guaranteeing robustness has long been an important design objective of control system analysis. *Stochastic robustness* is a simple numerical procedure that can be used to measure and gain insight into robustness properties associated with linear control systems. In the realm of aircraft control systems, problems such as the effects of flight condition perturbations and model-order uncertainties on robustness are easily and effectively analyzed using stochastic robustness. The concept of stochastic robustness will be reviewed and examples will be presented demonstrating its use in flight control system analysis.

- Summary of stochastic robustness
- Control system robustness with flight condition perturbations
- Control system robustness with model-order uncertainties

  Actuator dynamics
  Aeroelastic effects

- Summary of results

DEFINITIONS

Control system ***robustness*** is defined as the ability to maintain satisfactory stability or performance characteristics in the presence of all conceivable parameter variations. A good robustness measure is vital to guarantee and understand control system robustness. ***Stochastic robustness*** provides such a measure. It uses the statistics of a plant's variable parameters and Monte Carlo simulation to compute the probability distributions of closed-loop system characteristics. Present research has concentrated on stability robustness as characterized by the closed-loop eigenvalues, although the method can be extended to other closed-loop characteristics. Stochastic robustness is computationally simple. For a single Monte Carlo evaluation, random numbers are generated and shaped to match the parameter statistics, added to the mean parameter vector, and the closed-loop eigenvalues are computed using the modified parameters. Repeated Monte Carlo evaluations give rise to the ***stochastic root locus***, a plot of the probability distributions of the closed-loop eigenvalues. The ***probability of instability***, or probability that all of these eigenvalues lie in the open left-half s plane, is the scalar measure of robustness.

Robustness

The ability to maintain satisfactory stability/performance characteristics in the presence of all conceivable parameter variations.

Stochastic robustness

A robustness *measure* based on the probability distributions of closed-loop characteristics, given the statistics of a plant's variable parameters.

- Characteristics can be eigenvalues, performance, control authority, disturbance rejection
- Based on *Monte Carlo* simulation
- Not limited to Gaussian parameters

Stochastic root locus

Plot of the probability distributions of closed-loop eigenvalues.

Probability of instability

Probability that closed-loop system is unstable - a scalar measure of stability robustness.

STOCHASTIC ROBUSTNESS APPLIED TO DEMONSTRATOR AIRCRAFT

As an example of the application of stochastic robustness, three controllers are applied to a fourth-order longitudinal model of an open-loop unstable aircraft. A ten-element parameter vector consisting of elements of the dynamic and control effect matrices is chosen. The three control designs are chosen to reflect increasingly robust controllers. The first two Cases are LQR controllers with low and high control weighting respectively, and the third controller multiplies the Case (b) controller by a factor of five to restore the closed-loop bandwidth to that of Case (a). These three cases have been chosen not to satisfy any particular flying qualities criteria, but merely to demonstrate the impact of differing generalized design criteria on stochastic robustness.

Fourth-order longitudinal dynamic model

$$\dot{\mathbf{x}} = \mathbf{F(p)}\,\mathbf{x} + \mathbf{G(p)}\,\mathbf{u}$$

$$\mathbf{u} = -\mathbf{C}\,\mathbf{x}$$

Ten-element parameter vector

$$\mathbf{p} = [f_{11}\ f_{12}\ f_{13}\ f_{22}\ f_{32}\ f_{33}\ g_{11}\ g_{12}\ g_{31}\ g_{32}]$$

f_{ij}, g_{ij} are elements of matrices **F** and **G**.

Control design (C matrix) to demonstrate stochastic robustness

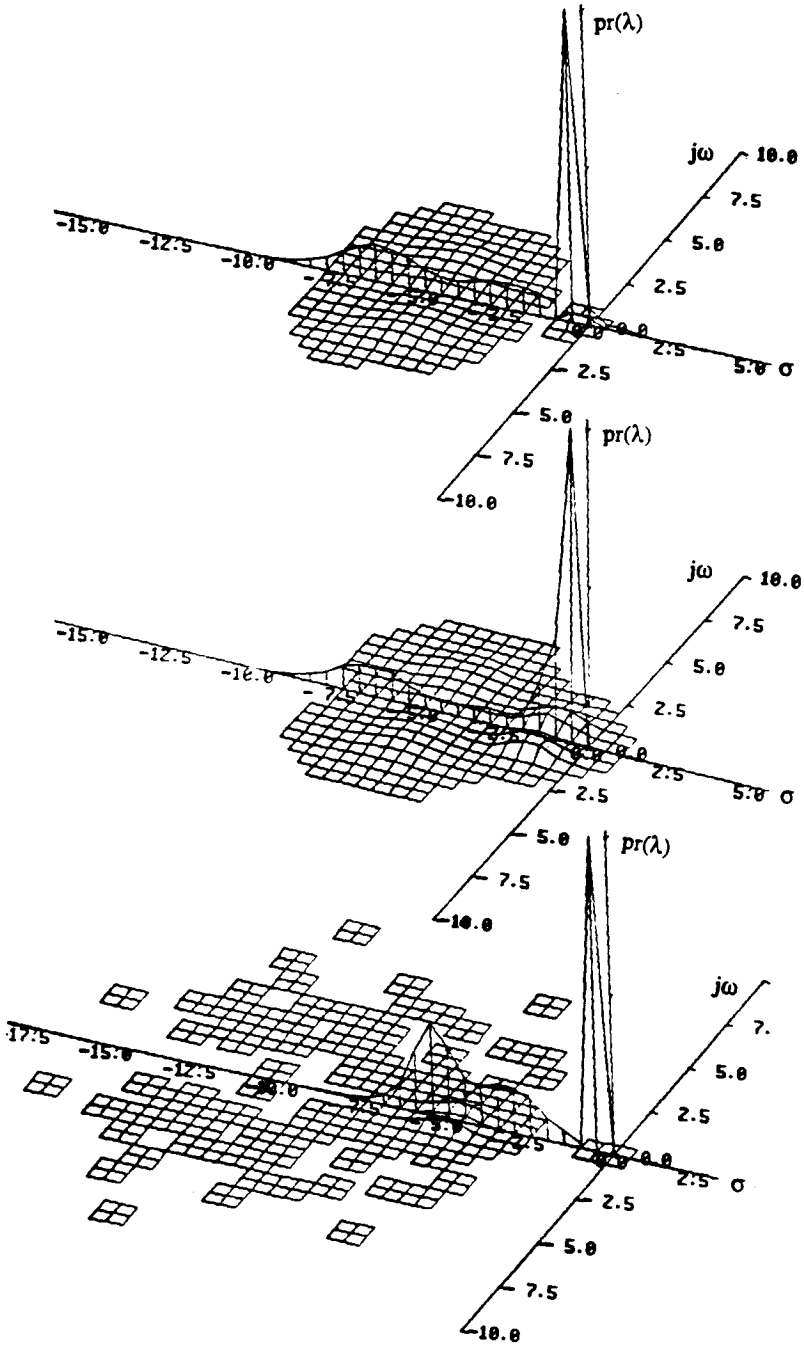
Case (a) LQR with low control weighting.

Case (b) LQR with high control weighting.

Case (c) Gain matrix of Case (b) is multiplied by 5 to restore bandwidth.

Stochastic root loci for demonstrator aircraft with 30% standard-deviation Gaussian parameters

The results for 30% Gaussian parameters and 10,000 Monte Carlo evaluations reflect the expected increase in robustness between control designs. The stochastic root locus shows the extent to which eigenvalues can vary. The eigenvalue near the origin is least affected by the parameter changes, and its peak dominates the distribution. In Cases (a) and (c), the left-most eigenvalue (not shown) has an enormous variance along the real axis. Interaction of roots around the origin causes instability. Robustness improves from Case (a) to (b) as control usage is restrained by high control weighting, and the ad hoc robustness recovery technique used in Case (c) gives additional improvement.



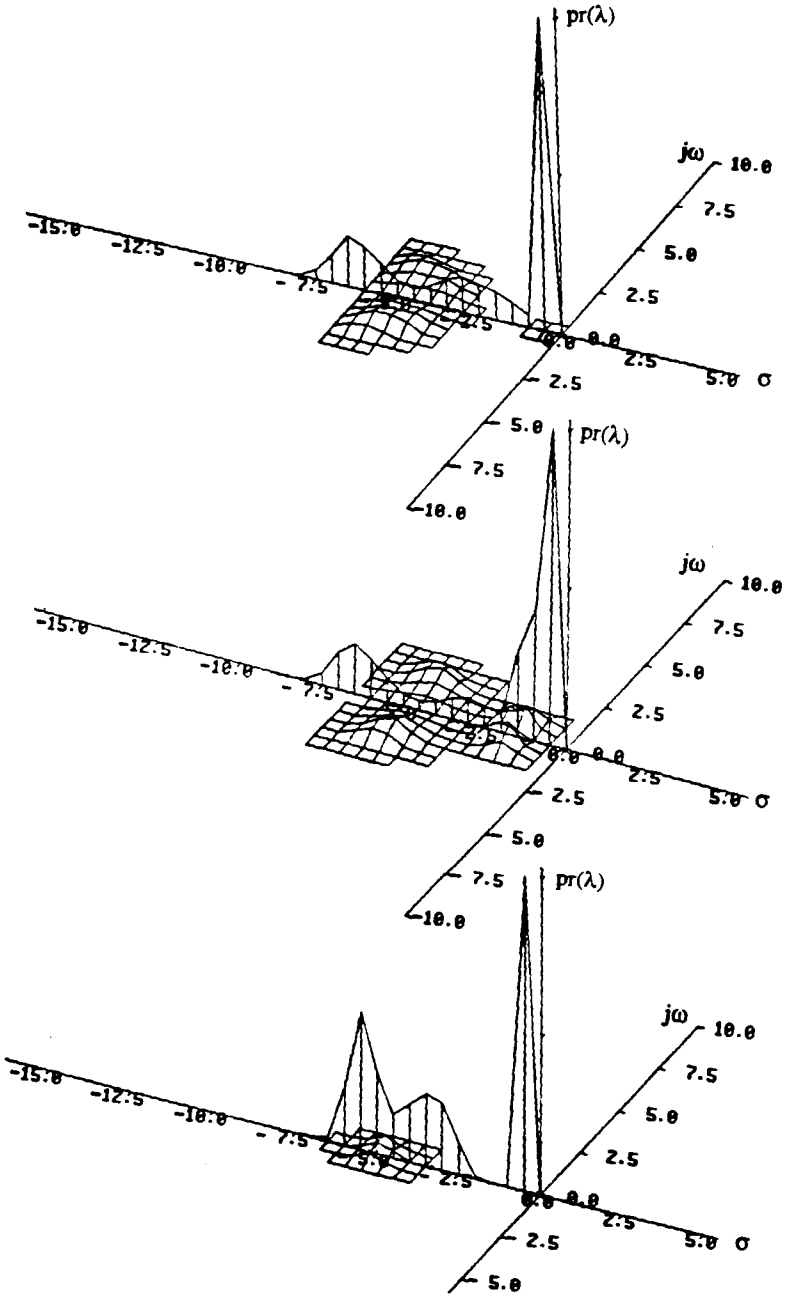
Case (a)
$\mathbb{P} = 0.0711$

Case (b)
$\mathbb{P} = 0.0169$

Case (c)
$\mathbb{P} = 0.0033$

Stochastic root loci for demonstrator aircraft with 30% uniformly distributed parameters

For 30% uniformly distributed parameters and 10,000 Monte Carlo evaluations, the probability of instability for all three cases is zero. The stochastic root locus gives a good indication of the effects of Gaussian "tails" on the eigenvalue probability densities.



Case (a)
$\mathbb{P} = 0.0$

Case (b)
$\mathbb{P} = 0.0$

Case (c)
$\mathbb{P} = 0.0$

CONTROL SYSTEM ROBUSTNESS WITH FLIGHT CONDITION PERTURBATIONS

Demonstrator Aircraft with Flight Condition Effects

Dynamic pressure variations can be considered separately and included in the parameter vector. Although velocity (V) and air-density (ρ) are essentially deterministic, including them as separate parameters gives the ability to look at flight condition perturbations around the nominal and eliminates correlation of the remaining parameters. A twelve-element parameter vector results. ρ and V are modeled as uniform parameters, giving an indication of stochastic robustness over a range of flight conditions.

Fourth-order longitudinal dynamic model

$$\dot{\mathbf{x}} = \mathbf{F}(\boldsymbol{p})\,\mathbf{x} + \mathbf{G}(\boldsymbol{p})\,\mathbf{u}$$

$$\mathbf{u} = -\mathbf{C}\,\mathbf{x}$$

Twelve-element parameter vector

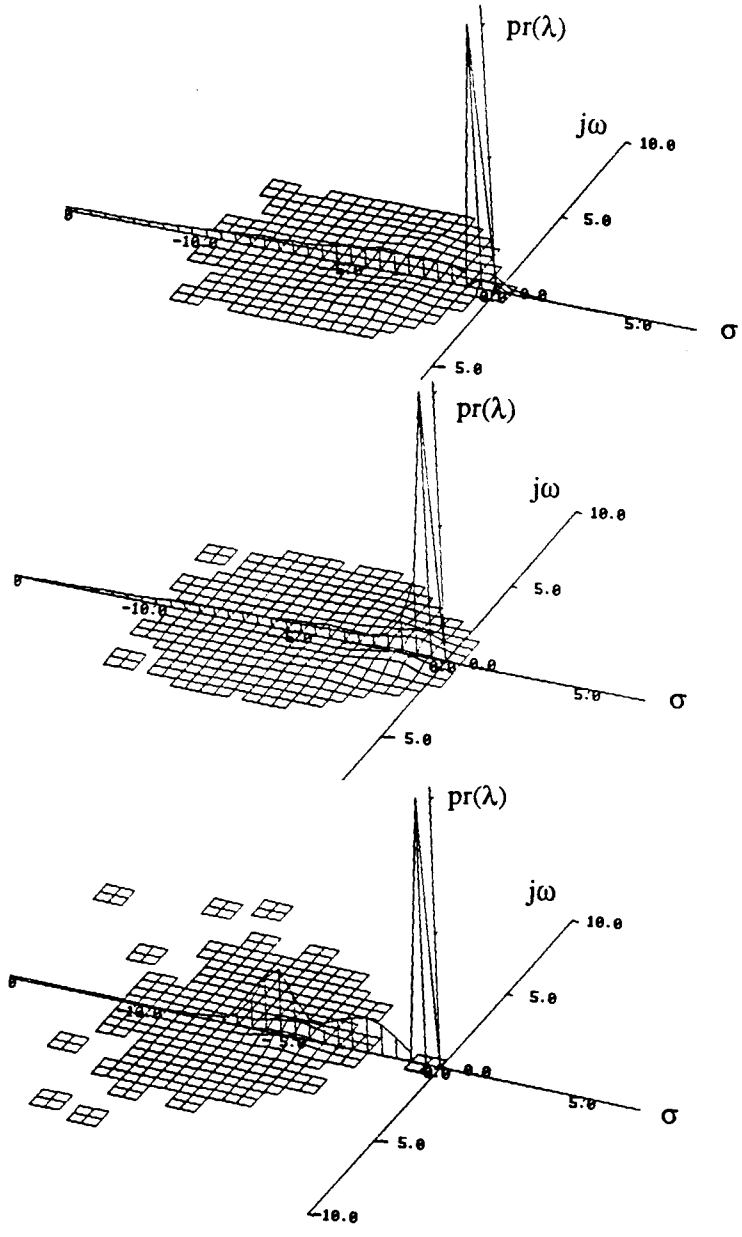
$$\boldsymbol{p} = [\rho\ V\ f_{11}\ f_{12}\ f_{13}\ f_{22}\ f_{32}\ f_{33}\ g_{11}\ g_{12}\ g_{31}\ g_{32}]$$

ρ is the air density (nominal value 0.00152 s/ft^3)
V is the velocity (nominal value 670 ft/sec)
f_{ij}, g_{ij} are elements of **F** and **G** with ρ and V considered separately.

Model ρ and V as uniform parameters and apply stochastic robustness using the same three control designs as previous example.

Stochastic root loci for demonstrator aircraft with 30% Gaussian parameters and 30% uniform ρ and V

Next, we examine the stochastic root loci for these three cases: for 30% uniform ρ and V, and 30%-standard-deviation Gaussian uncertainty on each of the remaining elements of the parameter vector. The shapes of the root loci are similar to the case with correlated parameters.



Case (a)
$\mathbb{P} = 0.0736$

Case (b)
$\mathbb{P} = 0.0166$

Case (c)
$\mathbb{P} = 0.006$

SUMMARY OF STABILITY ROBUSTNESS RESULTS FOR DEMONSTRATOR AIRCRAFT FOURTH-ORDER MODEL

Considering a case with non-varying flight condition along with the above results, the probabilities of instability seem to indicate that instability in Case (a) is a stronger function of uncertainties in individual parameters or stability derivatives than 30% V and ρ variations, while the remaining two cases are sensitive to flight condition variations.

	Controller		
	Case (a)	Case (b)	Case (c)
Correlated parameters 30% Gaussian variations	0.0711	0.0169	0.0033
Uncorrelated parameters 30% Gaussian variations 30% uniform ρ and V	0.0736	0.0166	0.0060
Uncorrelated parameters 30% Gaussian variations no ρ and V variations	0.0746	0.0162	0.0030

CONTROL SYSTEM ROBUSTNESS WITH ACTUATOR DYNAMICS

Stochastic robustness can be used to quantify the effects on robustness of actuator dynamics. First-order actuator dynamics are added for each control, resulting in a 14-element parameter vector. A controller is designed with LQR weighting specifications intended to approximate the controller of Case (a), while not pushing the actuator dynamics to unrealistic frequencies.

Fourth-order longitudinal dynamics and first-order actuator dynamics for each control

$$\dot{\mathbf{x}} = \mathbf{F}'(\boldsymbol{p}')\,\mathbf{x} + \mathbf{G}'(\boldsymbol{p}')\,\mathbf{u}$$

$$\mathbf{u} = -\mathbf{C}\,\mathbf{x}$$

14-element parameter vector

$$\boldsymbol{p}' = [\rho\ V\ f_{11}\ f_{12}\ f_{13}\ f_{22}\ f_{32}\ f_{33}\ g_{11}\ g_{12}\ g_{31}\ g_{32}\ \tau_c\ \tau_t\ ]$$

τ_c = canard time constant (nominal value 0.1 sec)

τ_t = thrust time constant (nominal value 1.0 sec)

- *Redesign Case (a) controller such that closed-loop longitudinal eigenvalues are the same as previous example and actuator dynamics are reasonable.*

- *Apply stochastic robustness using new controller.*

PROBABILITY OF INSTABILITY FOR VARIOUS GAUSSIAN CONTROL PARAMETER UNCERTAINTIES

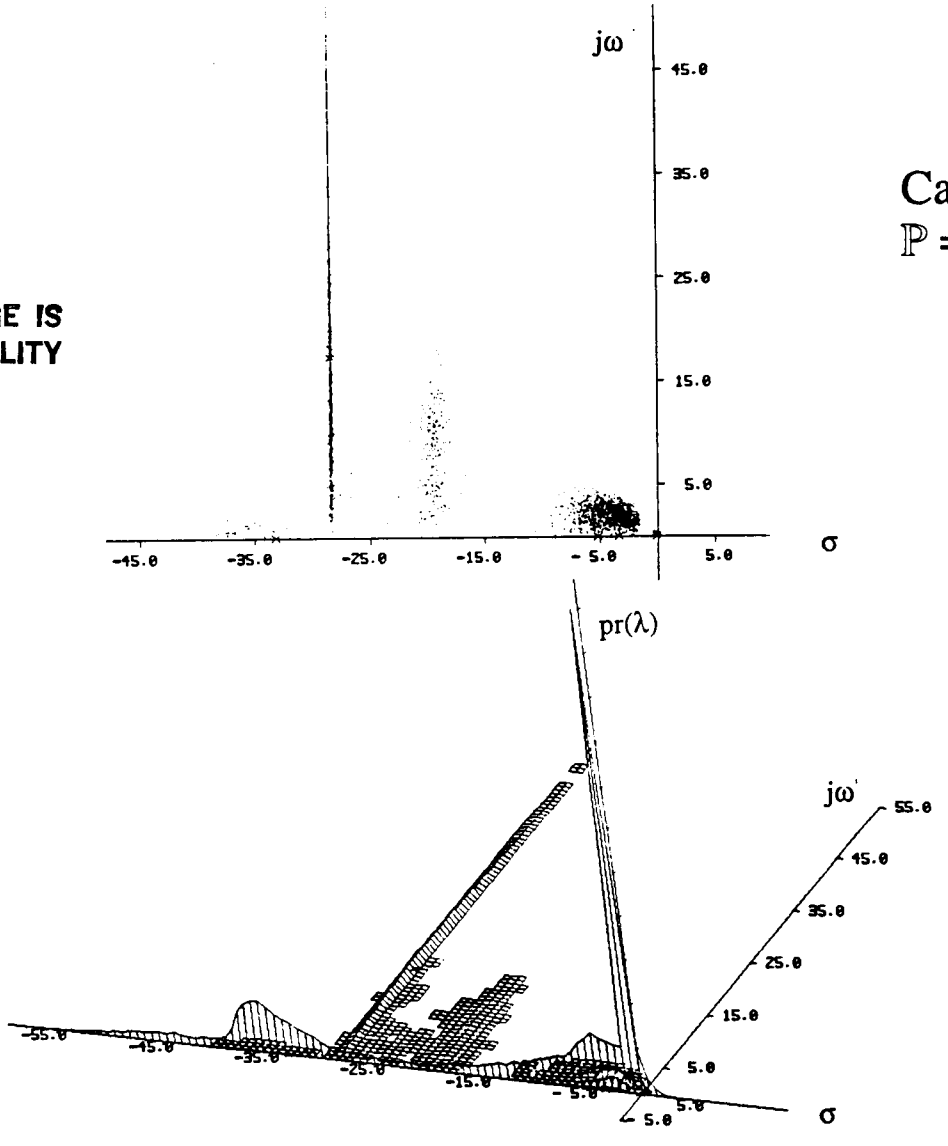
Stochastic robustness is applied for different values of the variance associated with each time constant, in order to detail the separate effects of each control lag. As indicated by the first line in the Table, simply including actuator dynamics increases the probability of instability, even if the associated parameters are known perfectly. This is a reasonable result because actuator dynamics are no longer infinitely fast but are allowed to interact with the rigid-body states. Qualitatively, bringing actuator dynamics in from infinity pushes the root-locus closer to instability. Stochastic robustness quantifies the effect. The thrust time constant has a small effect on the probability of instability, while a large increase in the probability of instability is seen as the canard time-constant standard-deviation increases from 30% to 150%.

ρ and V are 30% uniform parameters.
f_{ij}, g_{ij} are 30% Gaussian parameters.

standard-deviation of τ_c	standard-deviation of τ_t	$\mathbb{P}$
0	0	0.092
0	30	0.0988
30	30	0.101
30	150	0.1014
150	30	0.1474
No control dynamics		0.0736

Stochastic root loci for demonstrator aircraft with 30% Gaussian parameters, 30% uniform ρ and V, and non-varying τ_c and τ_t

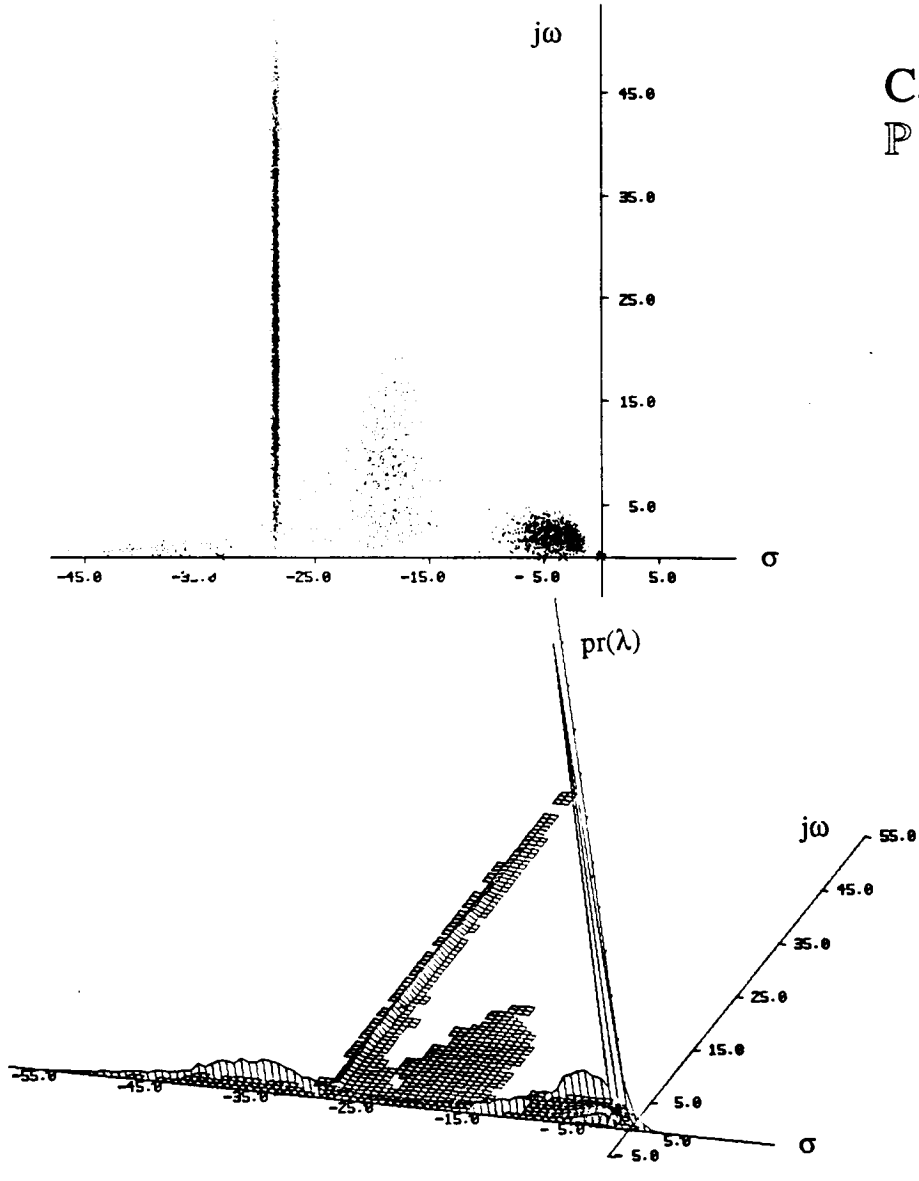
The stochastic root loci show that a strong coupling due to uncertainties can occur between the control and dynamic states, which tends to push more eigenvalues towards the right-half plane.



Case (a)
$\mathbb{P} = 0.092$

Stochastic root loci for demonstrator aircraft with 30% Gaussian parameters, 30% uniform ρ and V, and 30% Gaussian τ_c and τ_t

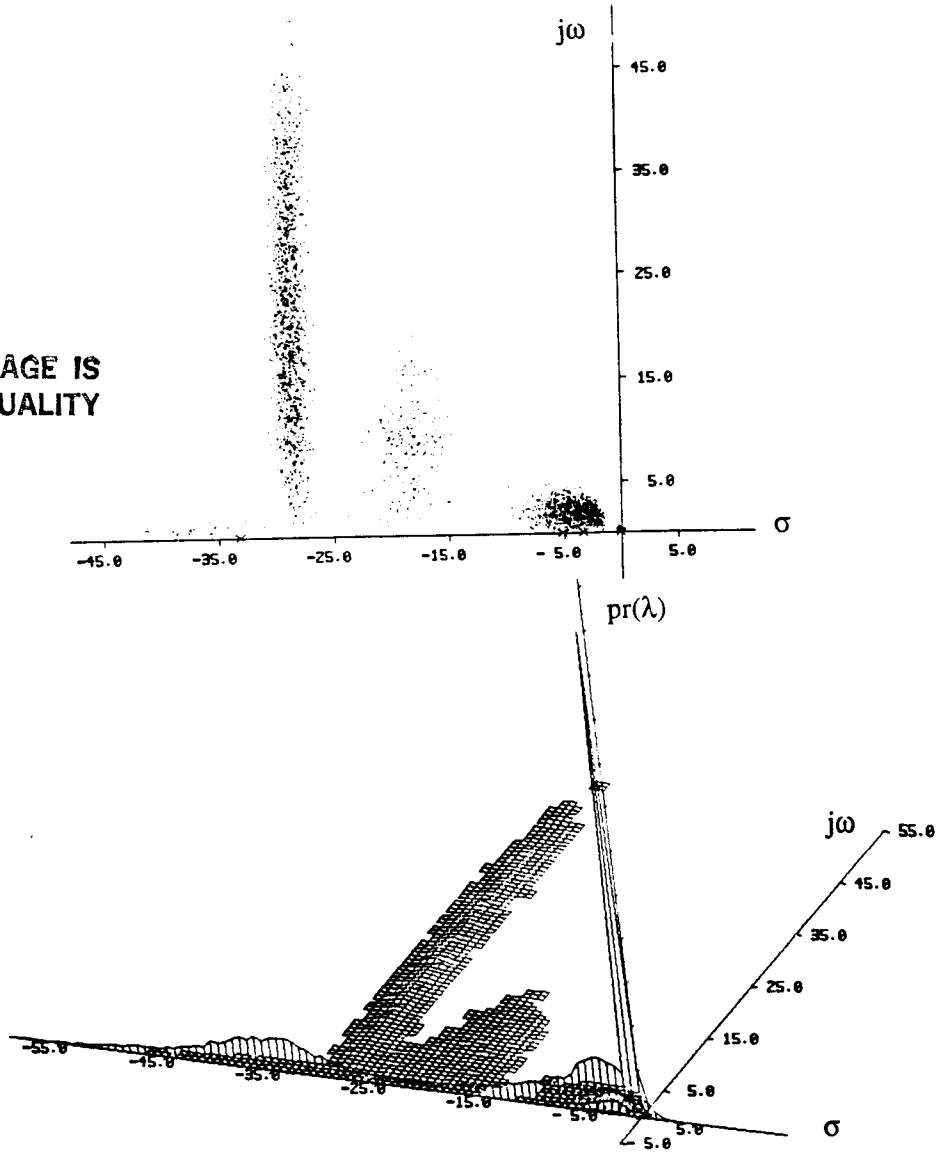
Increasing the standard-deviations associated with the time constants shows that the complex pair of eigenvalues has a small "variance" in the σ-direction, and a large variance in the $j\omega_d$ direction.



Case (a)
$\mathbb{P} = 0.101$

Stochastic root loci for demonstrator aircraft with 30% Gaussian parameters, 30% uniform ρ and V, 30% Gaussian τ_c, and 150% Gaussian τ_t

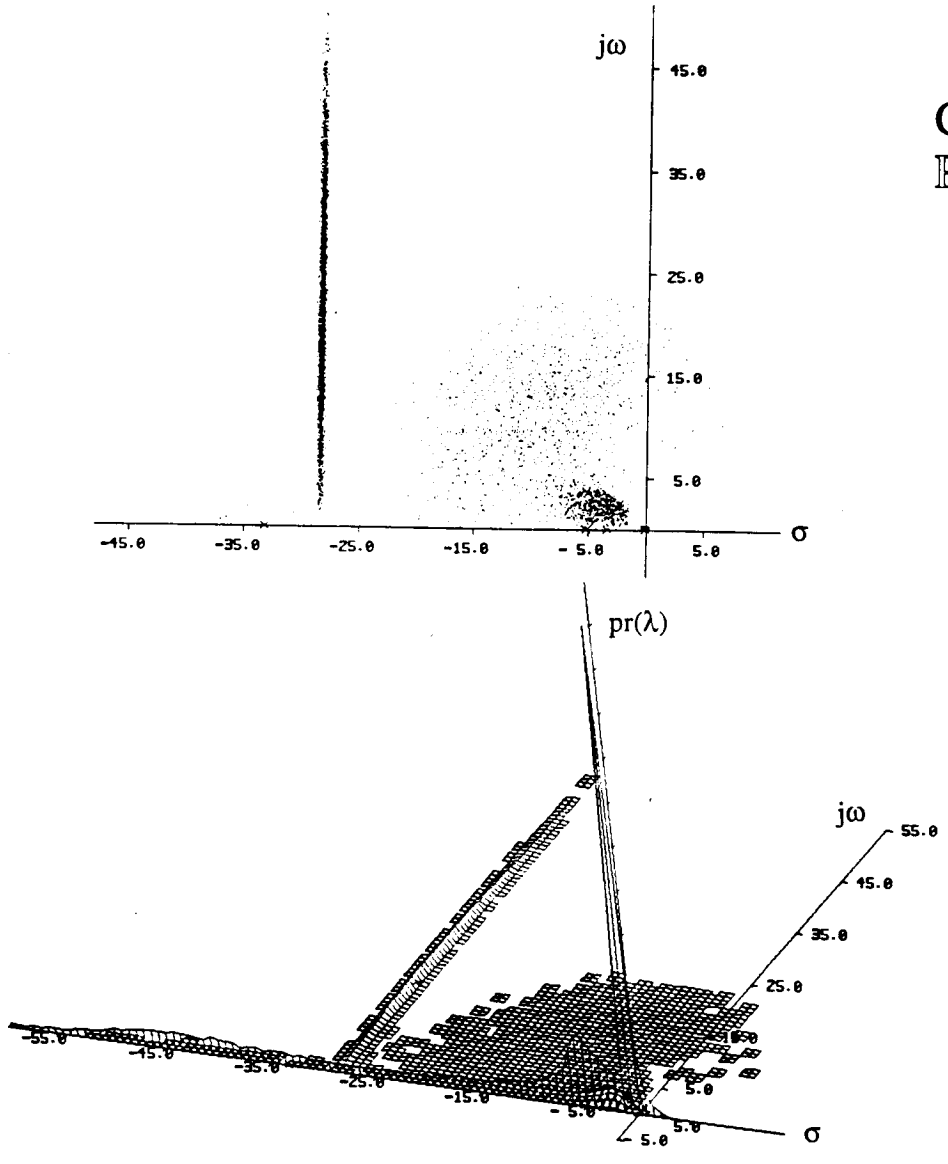
The σ-direction variance is largely due to the uncertainty associated with the thrust time constant, as illustrated by increasing the standard-deviation on this parameter to 150%. Increasing the uncertainty of τ_t has little effect on the probability of instability because it does not cause significant coupling with the dynamic modes.



Case (a)
$\mathbb{P} = 0.1014$

Stochastic root loci for demonstrator aircraft with 30% Gaussian parameters, 30% uniform ρ and V, 150% Gaussian τ_c, and 30% Gaussian τ_t

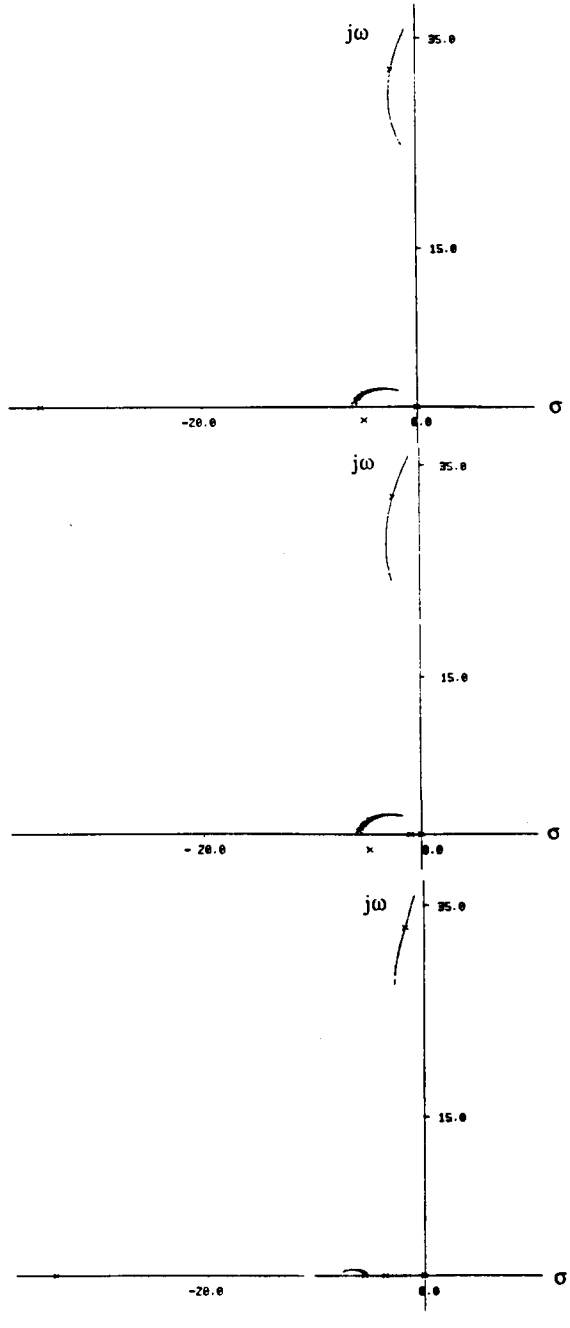
The variation at constant σ and coupling of controller and dynamic modes is largely due to variation of the canard time constant. Uncertainty in τ_c causes eigenvalues to migrate to the real axis and split off to form the complex "cloud" of eigenvalues that reaches instability.



Case (a)
$\mathbb{P} = 0.1474$

Stochastic root loci for demonstrator aircraft with aeroelastic effects for 30% uniform ρ and V

For 30% uniform variations in velocity and density alone, using the reduced-order gains, the probabilities of instability are zero. As expected, the closed-loop torsion eigenvalues at s= -0.1 ± 212.5j (not shown in figures) do not change with velocity and do not effect the probabilities of instability. Bending mode eigenvalues show a definite velocity trend, migrating towards instability as velocity increases. The closed-loop mean eigenvalues of rigid-body modes shift from the reduced-order case because of the presence of the added dynamics. In each case, the open-loop bending mode eigenvalues (-2.95 ± 32.03j) shift towards instability.



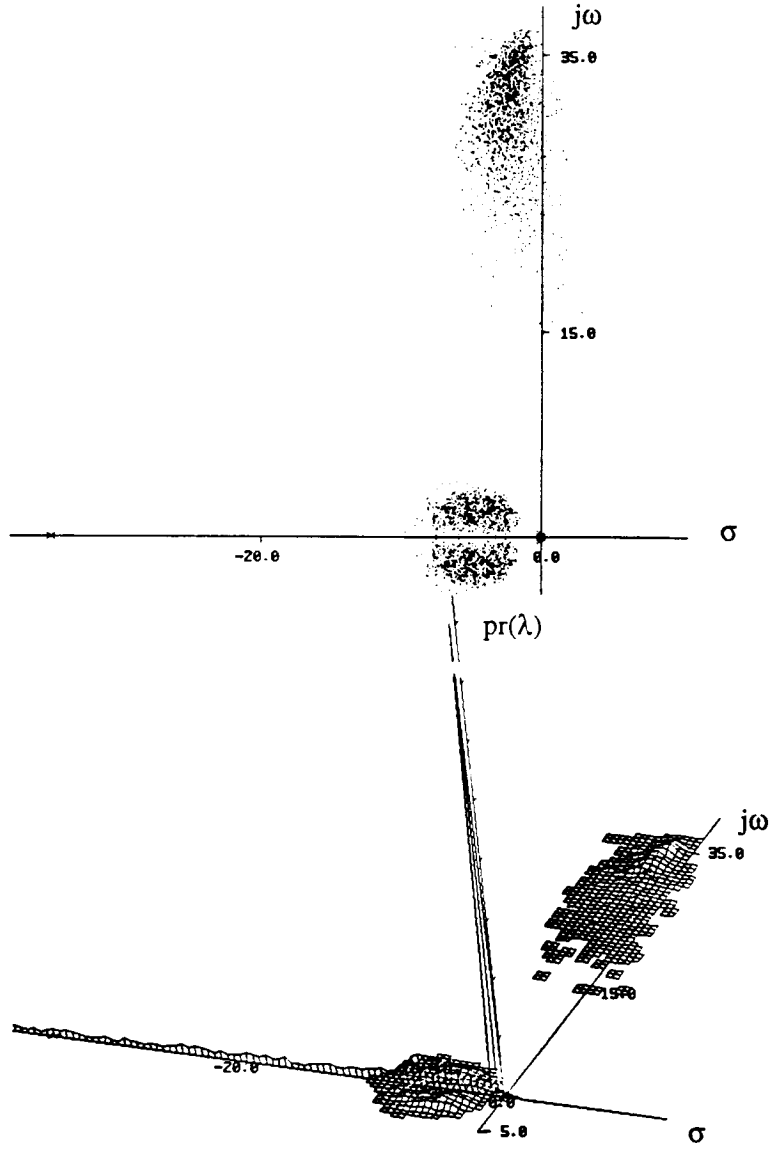
Case (a)
$\mathbb{P} = 0.0$

Case (b)
$\mathbb{P} = 0.0$

Case (c)
$\mathbb{P} = 0.0$

Stochastic root loci for demonstrator aircraft with aeroelastic effects for 30% Gaussian parameters and 30% uniform ρ and V

Next, the stochastic root loci for 30% uniform variations in ρ and V and 30% Gaussian variations of the parameters are evaluated. The peak near the origin is magnified to bring out the distribution associated with the bending mode eigenvalues.



Case (a)
$\mathbb{P} = 0.043$

CONTROL SYSTEM ROBUSTNESS WITH AEROELASTIC EFFECTS

With second-order, dynamic-pressure-dependent aeroelastic effects representing the wing's first bending and torsional modes, the (8x8) and (8x2) system matrices can be partitioned as shown, where $\mathbf{F_r} = \mathbf{F}$ and $\mathbf{G_r} = \mathbf{G}$ represent dynamic and control effects for rigid body modes, $\mathbf{F_{ra}}$ and $\mathbf{F_{ar}}$ couple the rigid and flexible modes, and $\mathbf{F_a}$, $\mathbf{G_a}$ represent aeroelastic dynamic and control effects. To separate effects of material properties, the four parameters K_b, M_b, K_t, and M_t are assumed to be known perfectly. The aeroelastic matrices introduce 32 additional parameters, and the resulting 44-element parameter vector includes separate ρ and V effects. For preliminary analysis, 40 parameters were used, although concern for statistical significance and limits on the computational facilities used to date calls for modification of the number of parameters for future studies.

Fourth-order longitudinal dynamics coupled with fourth-order aeroelastic effects

$$\dot{\mathbf{x}} = \mathbf{F}'(\boldsymbol{p}')\,\mathbf{x} + \mathbf{G}'(\boldsymbol{p}')\,\mathbf{u}$$

$$\mathbf{u} = -\mathbf{C}\,\mathbf{x}$$

$$\mathbf{F}' = \begin{bmatrix} \mathbf{F_r} & \mathbf{F_{ra}} \\ \mathbf{F_{ar}} & \mathbf{F_a} \end{bmatrix} \qquad \mathbf{G}' = \begin{bmatrix} \mathbf{G_r} \\ \mathbf{G_a} \end{bmatrix}$$

$\mathbf{F_r} = \mathbf{F}$, $\mathbf{G_r} = \mathbf{G}$ represent rigid body modes
$\mathbf{F_a}$, $\mathbf{G_a}$ represents coupled second-order bending and torsion modes
$\mathbf{F_{ra}}$ and $\mathbf{F_{ar}}$ couple rigid and aeroelastic modes

44-element parameter vector

$\boldsymbol{p}' = [\rho\ V\ f_{11}\ f_{12}\ f_{13}\ f_{22}\ f_{32}\ f_{33}\ g_{11}\ g_{12}\ g_{31}\ g_{32}$ + uncertain terms from $\mathbf{F_{ra}}$, $\mathbf{F_{ar}}$, $\mathbf{F_a}$, and $\mathbf{G_a}$]

CONTROL SYSTEM ROBUSTNESS WITH AEROELASTIC EFFECTS (CONTINUED)

In terms of structural-dimensional derivatives, $\mathbf{F_a}$ can be represented as given. Material properties dominate the torsional mode, which varies little with dynamic pressure variations. Stochastic robustness is applied to the new system using the reduced-order gains. Coupling of the systems through $\mathbf{F_{ra}}$ and $\mathbf{F_{ar}}$ causes the closed-loop system to be sensitive to aeroelastic terms.

• *Assume generalized mass and stiffness of bending and torsion modes are known perfectly.*

$$\mathbf{F_a} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ (S^1_{\eta_b} - K_b)/M_b & S^1_{\dot{\eta}_b}/M_b & S^1_{\eta_t}/M_b & S^1_{\dot{\eta}_t}/M_b \\ 0 & 0 & 0 & 1 \\ S^2_{\eta_b}/M_t & S^2_{\dot{\eta}_b}/M_t & (S^2_{\eta_t} - K_t)/M_t & S^2_{\dot{\eta}_t}/M_t \end{bmatrix}$$

K_b, M_b, K_t and M_t are generalized stiffness and mass for each mode.

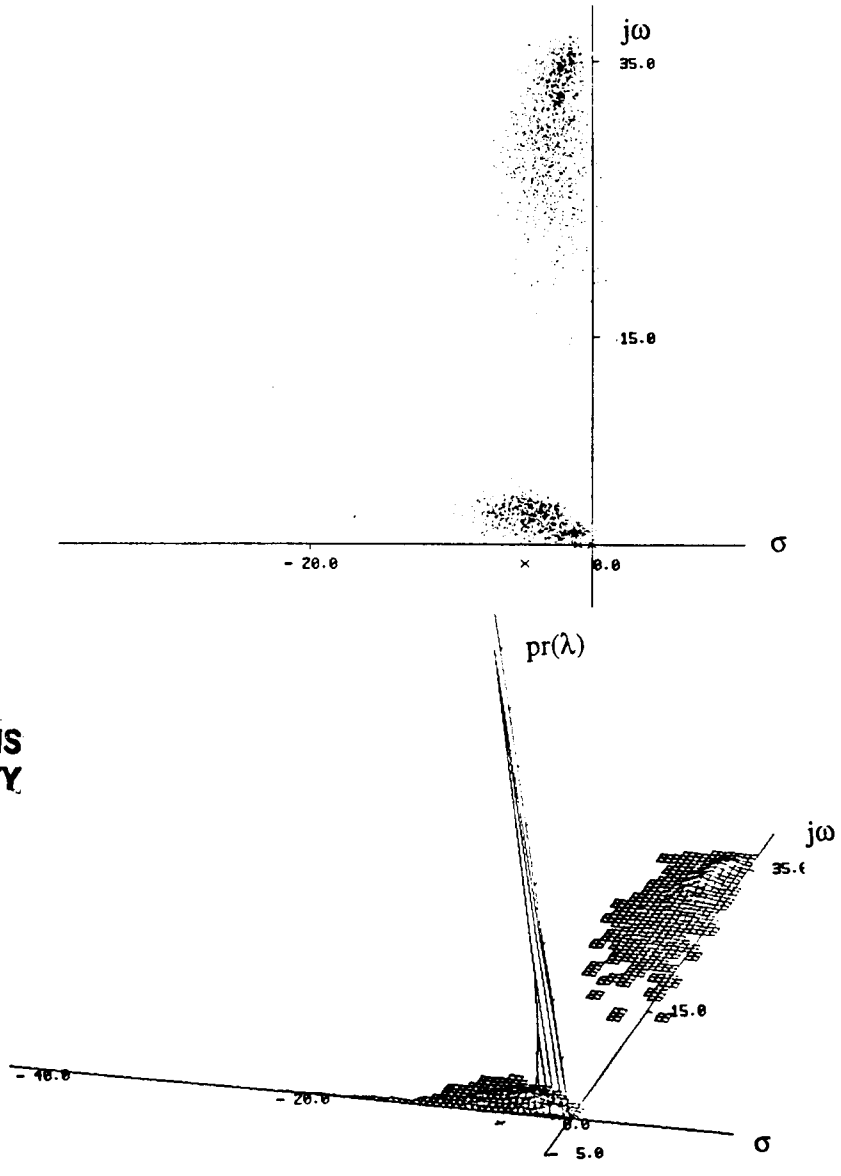
$S_\eta{}^i$ are structural dimensional derivatives.

• *Apply stochastic robustness to new system using gains established previously.*

$$\mathbf{C'} = [\, \mathbf{C} \;\; \mathbf{0} \,]$$

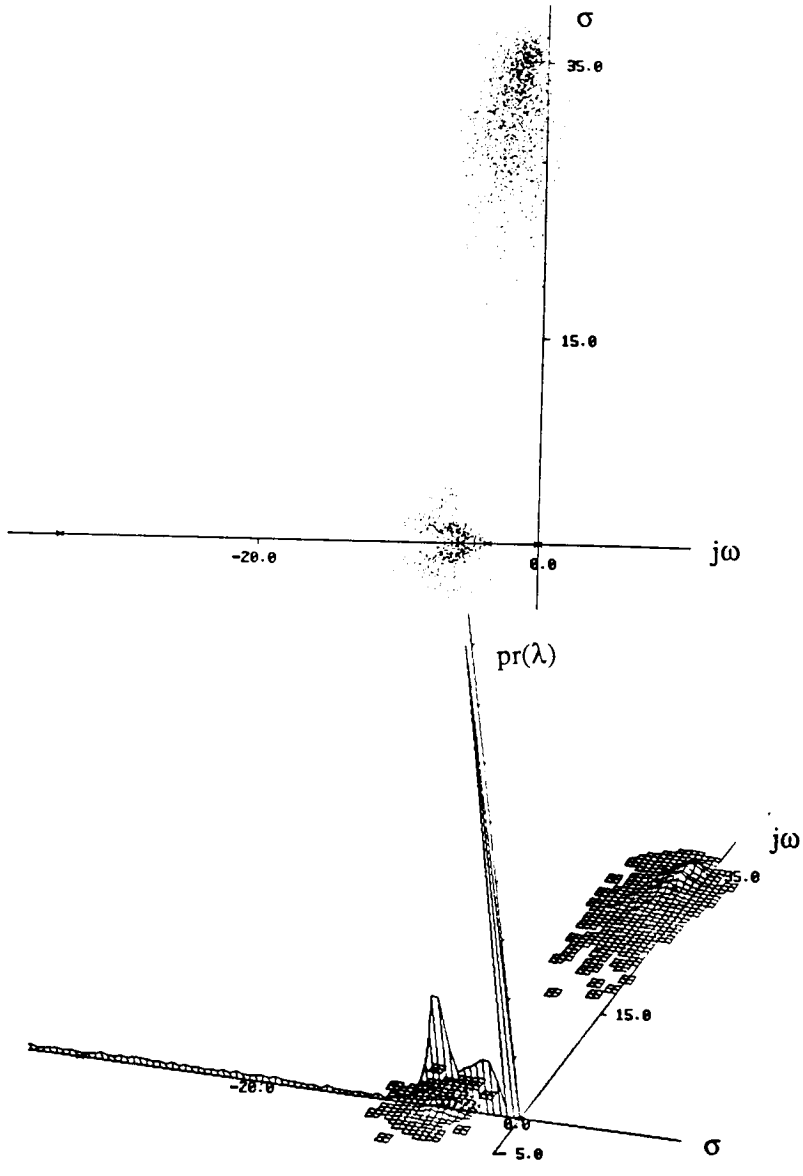
$$\mathbf{F_{closed\text{-}loop}} = \begin{bmatrix} \mathbf{F_r - G_r C} & \mathbf{F_{ra}} \\ \mathbf{F_{ar} - G_a C} & \mathbf{F_a} \end{bmatrix}$$

Stochastic root loci for demonstrator aircraft with 30% Gaussian parameters and 30% uniform ρ and V



Case (b)
$\mathbb{P} = 0.017$

Stochastic root loci for demonstrator aircraft with 30% Gaussian parameters and 30% uniform ρ and V



Case (c)
$\mathbb{P} = 0.0415$

SUMMARY OF ROBUSTNESS OF DEMONSTRATOR AIRCRAFT WITH AEROELASTIC EFFECTS

While definite conclusions cannot be reached because of the small sample space, this example illustrates the type of analysis possible using stochastic robustness. Certain trends are evident. The disparity in robustness between Cases (a) and (c) is reduced, and Case (c) shows a considerable decrease in robustness, while the robustness of the first two cases is at least retained or possibly improved. Application of a reduced-order controller to a higher order system does not guarantee that the robustness margins of the original system are retained, but the robustness of the system does not always go in the adverse direction. (This is somewhat analogous to the loss of guaranteed stability margins when applying LQG). Stochastic robustness again provides an excellent framework to quantify the effects of applying a reduced-order controller to a higher-order system. In each of these examples, questions concerning the selection of the number of Monte Carlo simulations, confidence limits, and statistical significance of results are issues of future research.

Probability of instability:

	without aeroelastic effects	with aeroelastic effects
Case (a)	0.0736	0.0430
Case (b)	0.0166	0.017
Case (c)	0.006	0.0415

Closed-loop eigenvalues:

	without aeroelastic effects	with aeroelastic effects
Case (a)	-0.02 -3.32, -5.14 -35.0	-0.02 $-4.96 \pm 1.27j$ -35.0 $-2.3 \pm 32.0j$
Case (b)	-0.02 -1.09 -3.36, -5.15	-0.02 -1.01 $-4.8 \pm 1.38j$ $-2.53 \pm 32.0j$
Case (c)	-0.01 -3.44, -5.15 -32.21	-0.02 -3.6, -5.53 -34.1 $-1.74 \pm 32.88j$

SUMMARY

Stochastic robustness offers a rigorous yet straightforward alternative to current metrics for control system robustness that is simple to compute and is unfettered by normally difficult problem statements, such as non-Gaussian statistics, products of parameter variations, and structured uncertainty. The approach answers the question, "How likely is the closed-loop system to fail, given limits of parameter uncertainty?" It makes good use of modern computational and graphic tools, and it is easily related to practical design considerations.

The examples presented here illustrate the use of stochastic robustness and its advantage in studying aircraft control systems. The parameters of aircraft stability and control effect matrices (stability derivatives and nominal flight condition parameters) lend themselves to this type of analysis tool. The stochastic robustness of different control system designs can be directly compared. Stochastic robustness can be used to study stability with flight condition variations. The method is also easily applied to model-order uncertainties in aircraft control systems by adding the uncertain dynamics to the system and assigning appropriate statistics to the new parameters. Quantitative effects of individual parameters or combinations of parameters on robustness can be measured in terms of the probability of instability. The principal difficulty in applying this method to control systems is that it is computationally intensive; however, requirements are well within the capabilities of existing computers. The principal advantage of the approach is that it is easily implemented, and results have direct bearing on engineering objectives.

1. Stochastic robustness can be used to study effects of flight condition perturbations on robustness.

- By considering flight condition parameters separately, parameters are uncorrelated.
- Can separate flight condition effects on robustness from parameter uncertainty effects.

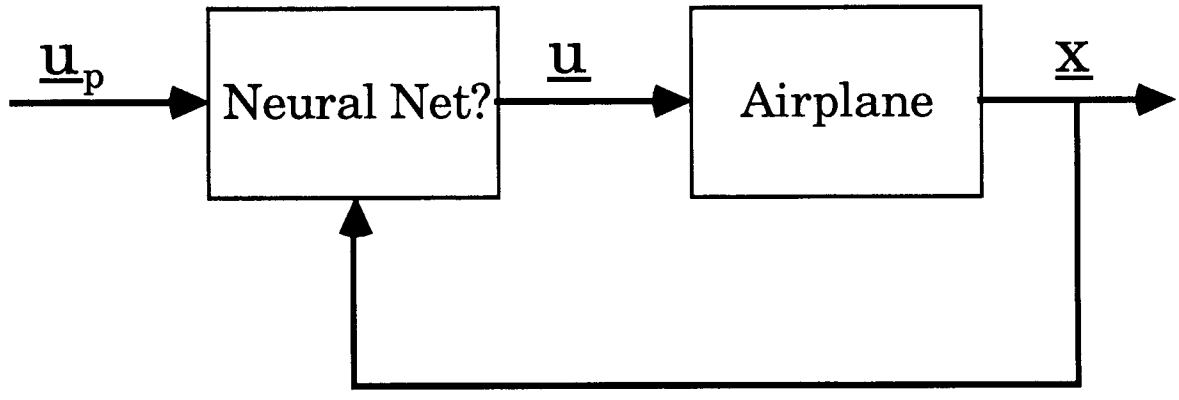
2. Stochastic robustness can be used to study effects of model-order uncertainties on robustness.

- Shows magnitude of actuator dynamics effect on robustness.
- Reveals instability or robustness degredation due to neglected dynamics.

Neural Networks for Aircraft Control

Dennis Linse
Department of Mechanical and Aerospace Engineering
Princeton University

Current research in Artificial Neural Networks indicates that networks offer some potential advantages in adaptation and fault tolerance. This research is directed at determining the possible applicability of neural networks to aircraft control. The first application will be to aircraft trim.



- Introduction to Neural Networks
- Neural Networks for Aircraft Trim
- Application of Neural Networks to Control

Introduction to Neural Networks

Artificial neural networks (usually called simply neural networks) are an attempt to model the processing behavior of the nervous system. Neural network research focuses on a much larger group of networks loosely classified by the dense interconnection of simple computational elements. Since the majority of the processing operations are independent of one another, neural networks can conduct massively parallel computations. While the human brain is estimated to have 100 billion neurons each with approximately 1000 inputs and outputs, artificial neural networks are interested in the computational capabilities of smaller networks which may or may not be biologically correct.

- Artificial neural networks

  Abstract simulation of real nervous system

  Dense interconnection
  of simple computational elements

  Massively parallel computation

  Biologically inspired vs. biologically accurate

  Estimated 100 billion (10^{11}) neurons in brain
  Each has 1000 inputs and outputs

Introduction to Neural Networks

Most current applications of neural networks have been in the area of pattern recognition. A smaller number of applications have been devoted to optimization, most notably the famous Traveling Salesman Problem in which the shortest route through selected cities is determined.

The major benefits of neural network are two-fold. First, neural networks have the ability to learn their internal knowledge from presentation of input-output data. This means that the network does not have to be programmed in the traditional sense. It must be trained with examples of the desired input-output relationship. Adaptability arises if the learning process continues while the network is in operation. Neural networks are potentially very fault tolerant due to the massively parallel architecture. The knowledge contained in the network is distributed throughout the network so that the loss of individual computational elements should not seriously degrade the performance of the network.

- Applications

  Speech/Image/Pattern Recognition, Classification and Restoration

  Optimization

- A few potential benefits

  Adaptability/Learning

  Robustness / Fault tolerance

Introduction to Neural Networks

Neural network models are differentiated by the type of computational element (or node), the connection of the nodes, and the learning rule used to update the weighted connections between nodes.

- Neural network models are specified by

  Node characteristics

  - linear / nonlinear
  - analog / discrete (binary)

  Net topology

  - node interconnections
    - layers
    - feedforward/feedback

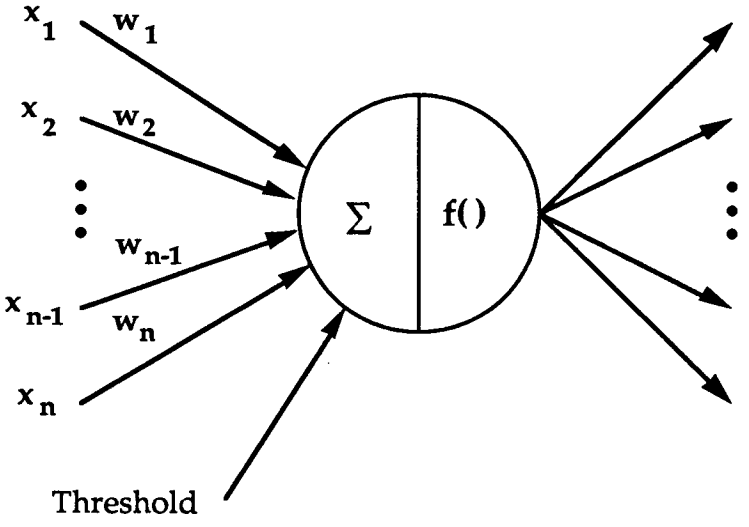
  Learning rule

  - supervised / unsupervised

Neural Network Node Characteristics

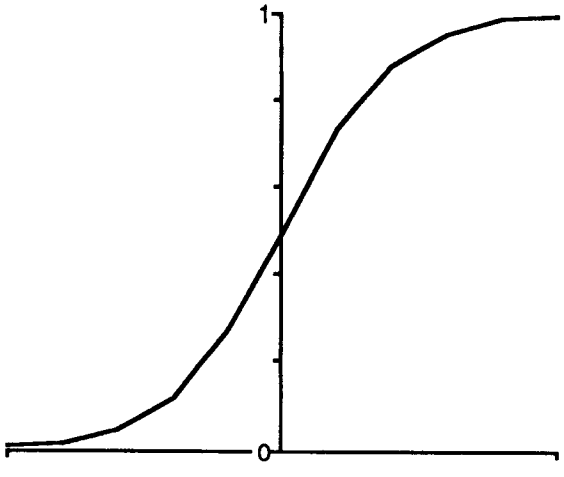
An individual neural network node computes a weighted sum of its inputs and runs the sum through a fixed function. The resulting value is transmitted to all of the following network elements. The fixed function, called the node activation function, is generally a bounded, S-shaped non-linear element. A common activation function is the sigmoid function. The nonlinearity of the activation function provides the computational power of the network. Networks with linear elements have been tested, but most contain some type of nonlinear function.

- Individual Neural Net Node



- Node Activation Function

Sigmoid $f(x) = 1.0 / (1.0 + \exp(-x))$

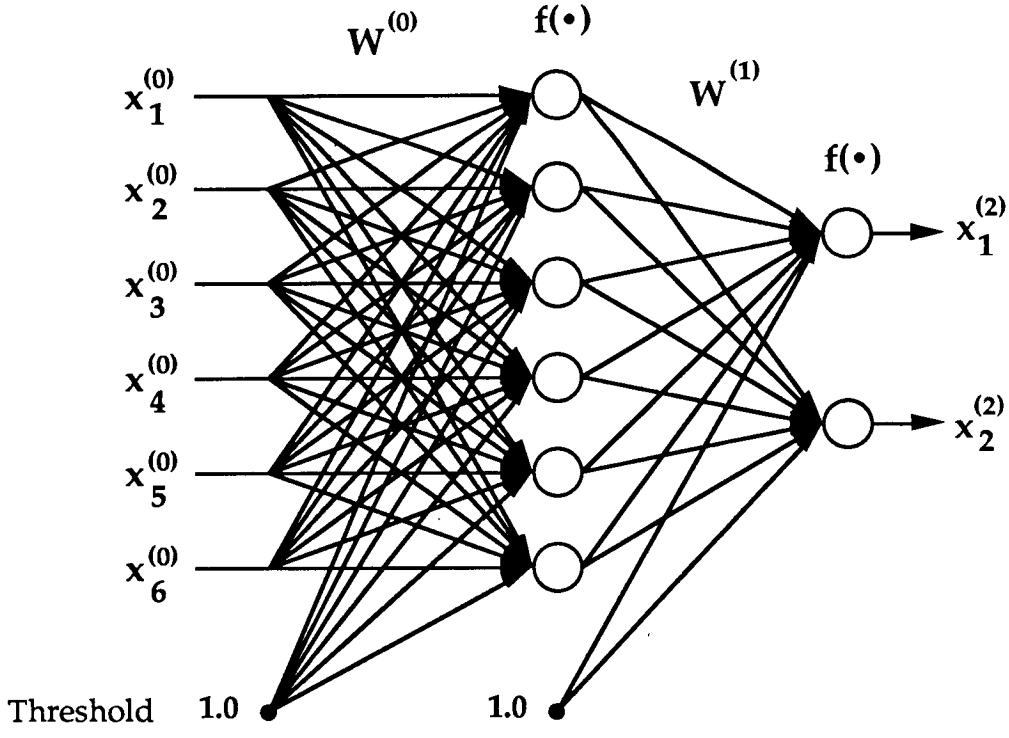


Network Topology and Operation

The multilayer feedforward network is characterized by the distinct layers of network nodes connected only in the forward direction. There are no connections between nodes in a single layer nor are there connections from higher layers to lower layers. A constant unit input is used as a threshold or bias.

The operation of the network is easily characterized by a recursive matrix equation. The output of each layer is simply the weighted sum of its inputs passed through the nonlinear activation function f. The knowledge of the system is contained in the weighted connections, i.e., the weights, W. As the weights change, the nonlinear input-output relationship function modeled by the network changes.

- Multilayer Feedforward Network



- Forward Operation (Activation, Retrieval)

$$\underline{x}^{(k)} = \underline{f}\left(\mathbf{W}^{(k-1)} \underline{x}^{(k-1)} \right), \qquad k = 1, \ldots, N \text{ layers}$$

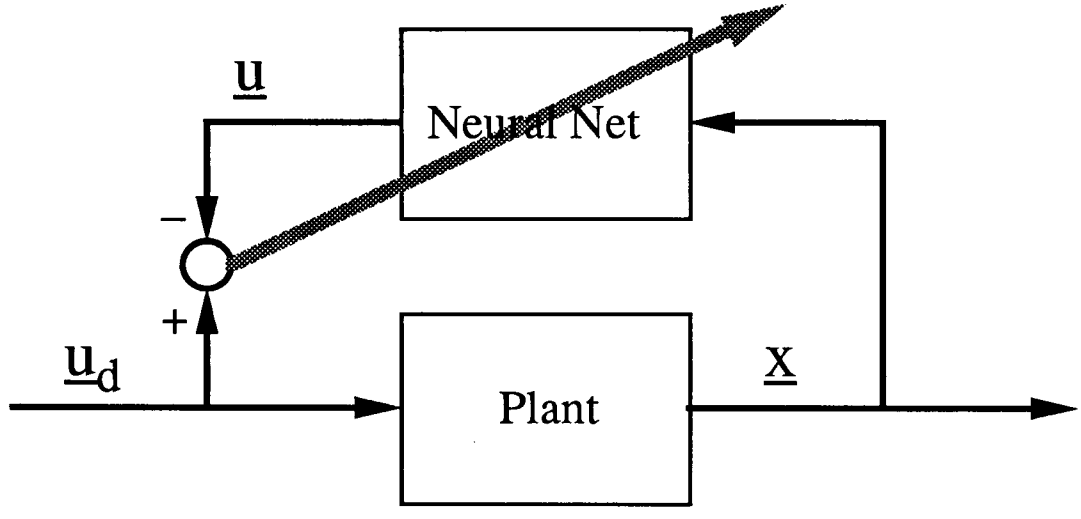
$$\underline{x}^{(0)} = \text{input}$$

$$\underline{x}^{(N)} = \text{output}$$

Neural Network Learning

In supervised learning, the neural network is trained to match a particular input-output relationship. First an input is applied to the neural network. Next, the output produced by the network is compared to the desired output. Finally, the network weights are adjusted based upon the error.

- Supervised Learning Diagram



Neural Network Learning

The backpropagation learning algorithm adjusts the weights of the network to minimize the mean square error of all of the outputs. The difficulty comes when assessing the importance of individual weights of the internal (or hidden) layers on each output. The backpropagation learning algorithm explicitly assigns a portion of the error to each element. The one requirement is that the nonlinear function in each network node must be differentiable.

- Learning by Backpropagation

  Adjust weights to minimize the mean square error

  Difficulty of error assignment

  Differentiability of nonlinearity

- Backpropagation

  $\mathbf{W}_{i+1} = \mathbf{W}_i + \Delta\mathbf{W}_i$

  $\Delta\mathbf{W}_i = \beta\, \underline{x}^{(k-1)}\, \underline{\delta}^{(k)}$

  $\underline{\delta}$ is the error term

  for output layer = $(\underline{u}_d - \underline{u})$

  for hidden layers, error is backpropagated

Neural Network for Aircraft Trim

A multilayer feedforward neural network is trained on the input-output relationship for the longitudinal trim of a transport aircraft. Using the backpropagation learning rule, the network is taught to produce the aircraft control positions necessary to maintain a given trim state. Such a trim map could be used as part of a perturbation control law. The power of the neural network would eventually be in the ability to continuously update its knowledge of the aircraft during operation.

- Exploit learning/adaptability features of neural networks

- Teach neural network the map between
  trim states and trim control positions

  Useful for

  Autotrim system
  Trim map for perturbation control law

- Straight and level flight
  Longitudinal states and controls for learning

$\underline{x}$ = { $\bar{q}$ - dynamic pressure

V - forward velocity

γ - flight path angle

α - angle of attack

q - pitch rate

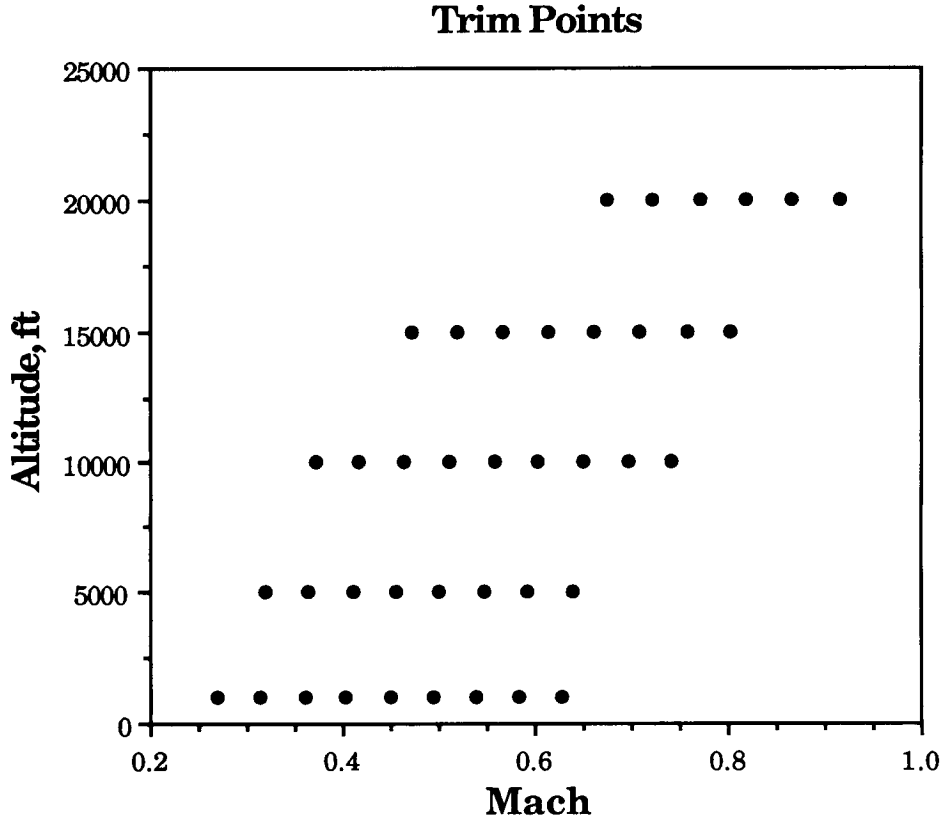
h - altitude }

$\underline{u}$ = { δ_E - elevator deflection

δ_T - throttle deflection }

Aircraft Trim Points

The neural network is trained on aircraft trim points ranging from Mach 0.25 at 1000 feet up to Mach 0.9 at 20000 feet.



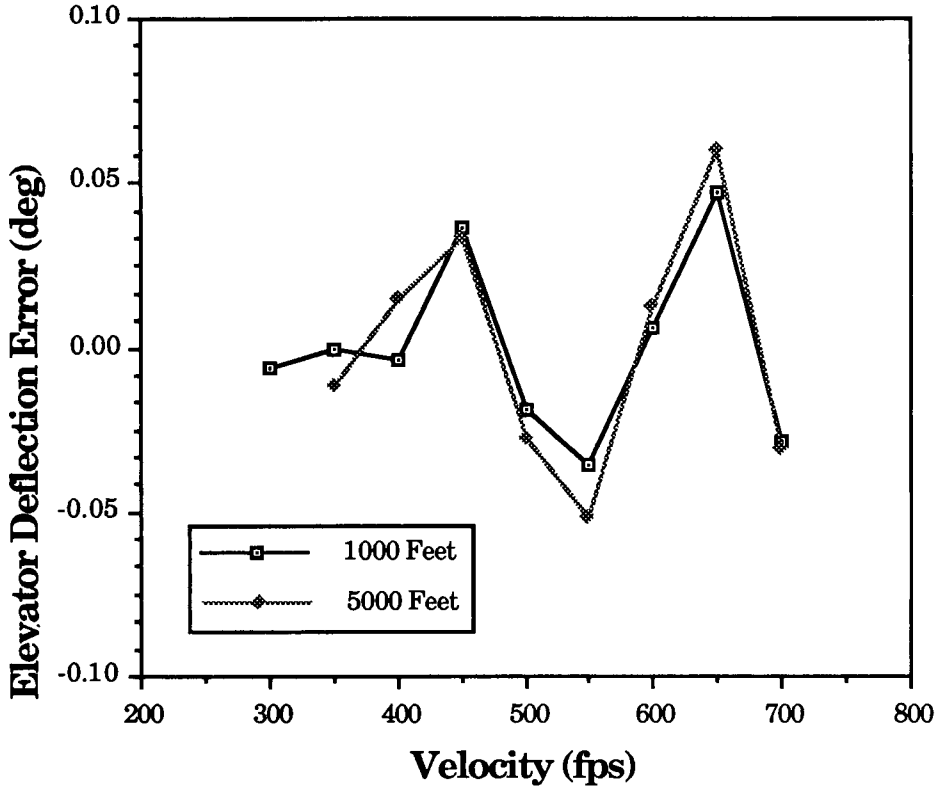
Example Network Run

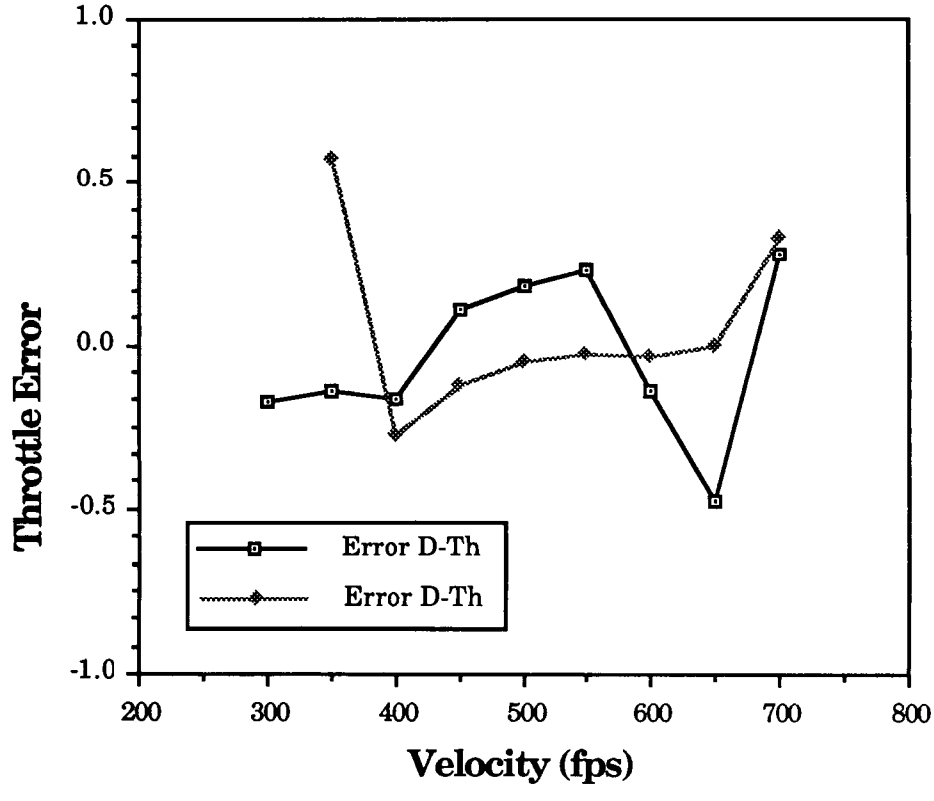
A simple example is shown in which the network is trained only on the aircraft data at 1000 and 5000 foot altitudes. A three layer network is used with 6 inputs, 6 hidden nodes in one layer, and 2 outputs.

- Train network with 17 trim points, 6 units in hidden layer
- 5000 Iterations, $\beta = 0.5$, momentum = 0.9
- Final Total Square Error = 0.0004

Error in Learned Response

The results after 5000 iterations are quite good. The elevator error is less than 1 percent and the throttle error is less than 3 percent over the entire learned range.

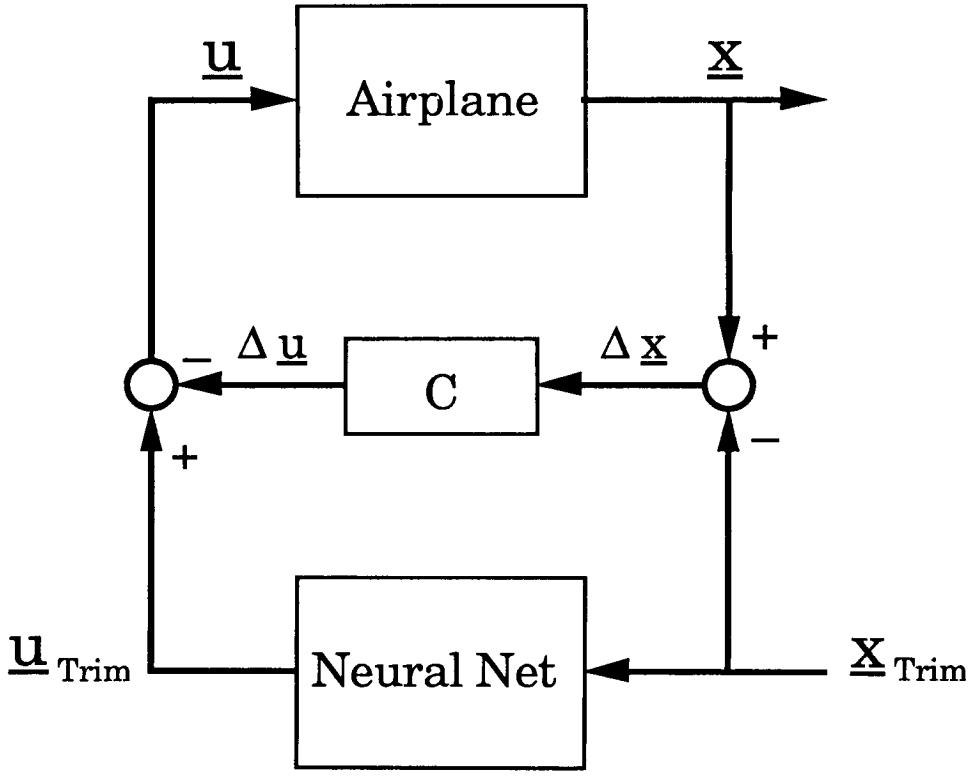




Neighboring Optimal Control Law using Neural Network

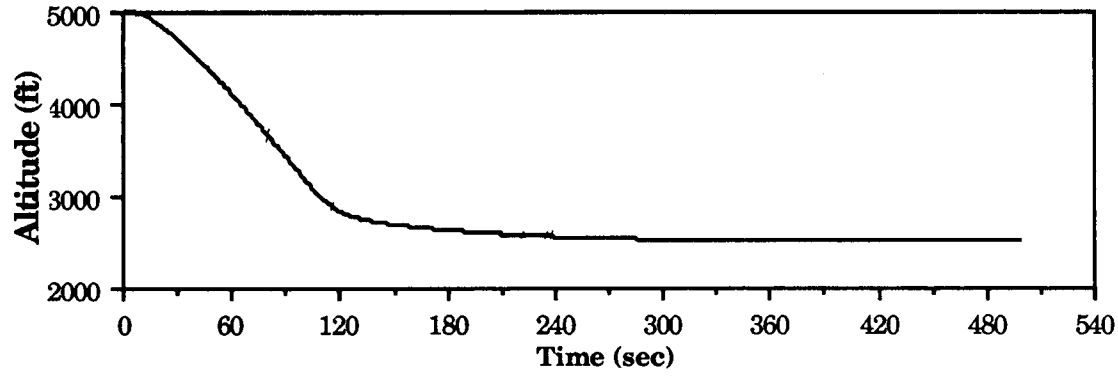
The neural network trained as a trim map fits quite nicely into a neighboring optimal control scheme. A perturbation control law is used to eliminate the errors between the actual aircraft state and the desired trim value. The resulting perturbation control position is combined with the trim control position from the neural network trim map to produce the control input for the aircraft.

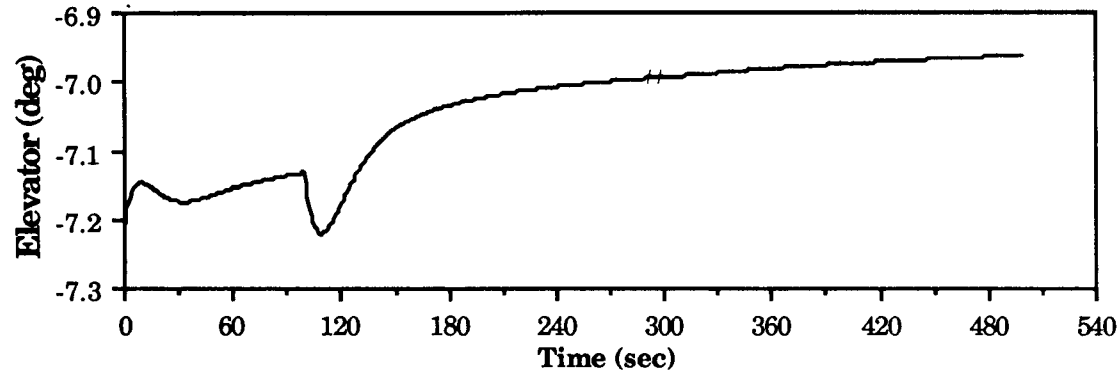
- Full state feedback perturbation control law developed for linearized system
- Applied to full nonlinear system using a Neighboring Optimal Control structure
- Trim values provided by Neural Network

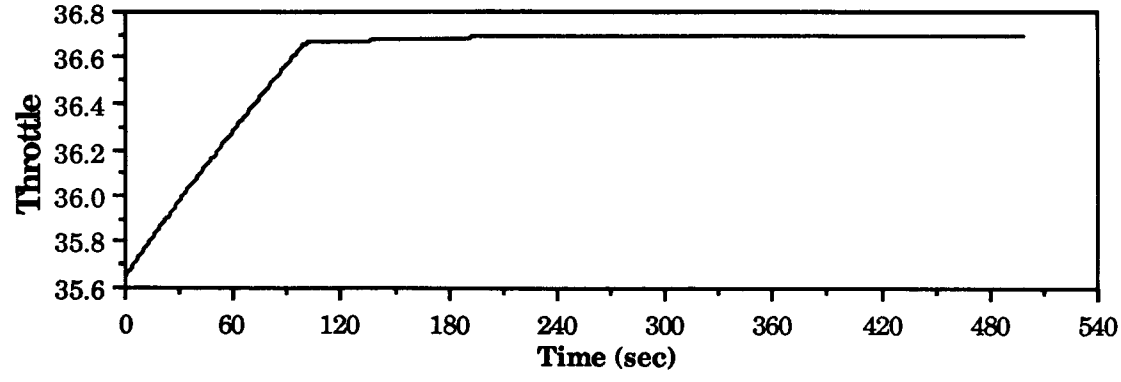


Example Histories using Neighboring Optimal Control with a Neural Net

Using the neighboring optimal control scheme with a neural network as the trim map, a smooth transition from 5000 to 2500 feet is effected for the simulated aircraft. The control positions are smooth and accurate over the entire range.



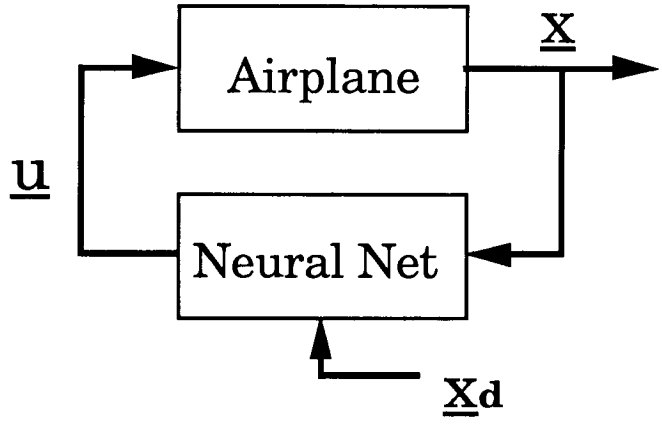




Neural Networks for Control

In addition to use as a trim map, a neural network could possibly be trained as a feedback control element. If the aircraft is modeled by a discrete time linear system, the weights could be adjusted as to minimize the error between the actual system states and some desired states. The mathematics of such a minimization appears feasible.

- Investigate Neural Network as Feedback Control Element



- Discrete Time Linear System

$$\underline{x}_{i+1} = \Phi\, \underline{x}_i + \Gamma\, \underline{u}_i$$

- Neural Network Control Law

$$\underline{u}_i = \underline{f}(\, W^{(1)}\, \underline{f}(\, W^{(0)}\, \underline{x}_i)\,)$$

- Adjust weights to minimize least squares fit error

$$I = \tfrac{1}{2} \sum_{i=1}^{M} (\underline{x}_d(i) - \underline{x}(i)\,)^T (\underline{x}_d(i) - \underline{x}(i)\,)$$

- Steepest descent training based on error gradient

$$W^{(l)}_{ij_{k+1}} = W^{(l)}_{ij_k} - \beta \frac{\partial I}{\partial W^{(l)}_{ij_k}}$$

Neural Networks for Aircraft Control

There are many topics which still must be investigated before any valid conclusions may be made about the usefulness of neural networks in aircraft control. The ability of the network to correctly generalize to trim points between the trained data is an important topic which must be looked into. The fault tolerance of neural networks is based upon massive parallelism. It is unknown how much of this fault tolerance is retained by smaller networks. There are many more possible applications for neural networks that are, as yet, untapped.

Topics of Interest for Neural Networks

- Generalization capabilities of networks

  What happens between learned points?

- Fault tolerance capabilities of networks

  What happens if nodes or weights malfunction?

- Other uses of neural networks for aircraft control

  Model identifier

  Adaptive control element

Conclusions

- Neural Networks show some promise for control applications
- There is much left to investigate

An Expert System for Wind Shear Avoidance

Robert F. Stengel[1] and D. Alexander Stratton[2]
Princeton University
Department of Mechanical and Aerospace Engineering
Princeton, NJ

ABSTRACT

A study of intelligent guidance and control concepts for protecting against the adverse effects of wind shear during aircraft takeoffs and landings is being conducted, with current emphasis on developing an expert system for wind shear avoidance. Principal objectives are to develop methods for assessing the likelihood of wind shear encounter (based on real-time information in the cockpit), for deciding what flight path to pursue (e.g., takeoff abort, landing go-around, or normal climbout or glide slope), and for using the aircraft's full potential for combating wind shear. This study requires the definition of both deterministic and statistical techniques for fusing internal and external information, for making "go/no-go"decisions, and for generating commands to the aircraft's autopilot and flight directors for both automatic and manually controlled flight. The program has begun with the development of the WindShear Safety Advisor, an expert system for pilot aiding that is based on the FAA Windshear Training Aid; a two-volume manual that presents an overview, pilot guide, training program, and substantiating data provides guidelines for this initial development. The WindShear Safety Advisor expert system currently contains over 200 rules and is coded in the LISP programming language.

INTRODUCTION

Flight in strong wind shears, especially microbursts, poses a unique and severe hazard to aircraft. The disturbance caused by the wind field may literally exceed the performance characteristics of the aircraft, making safe transit impossible even with optimal guidance and control strategies. An unusual degree of piloting skill may be required to successfully elude danger. Nevertheless, planes fly in moderate wind shear all the time; pilots learn to handle crosswinds, gustiness, and moderate frontal activity. The problem is that microbursts are random, rare phenomena; pilots do not develop the needed skills for coping with wind shear through normal experience. The typical pilot is likely to be confronted with a life threatening wind shear only once or twice during an entire flying career; hence, it is unlikely that he or she can learn all the important signs of wind shear and maintain a high level of proficiency in the proper control procedures.

On-board computation provides an excellent opportunity to assist the pilot in surviving encounters with severe wind shears, but the logic that must be executed in real time is complex and must have sufficient inputs for framing decisions about appropriate control actions. The computer program(s) and hardware to perform this task must have attributes of expert systems and control systems, they must account for the limitations of aircraft performance, and they must operate in real time. At least as important as its technical specifications, the on-board system must provide a satisfactory interface with the flight crew, which bears the ultimate responsibility for assuring safety. This means not only that the system must deduce near-optimal strategies and tactics for emergency situations but that it must distinguish between truly hazardous conditions and the more likely alternatives associated with normal aircraft operations.

Background on Wind Shear Encounter

The microburst phenomenon can be visualized as a vertically descending column of air that spreads out upon hitting the ground [1,2]. In a typical scenario, the aircraft first experiences a headwind as it enters the outflow, causing the aircraft to balloon above the flight path if no corrective action is taken (Fig. 1). (This phase may be preceded by an encounter with upflow and tailwinds at the edge of a ring vortex.) Just as the pilot is throttling back to accommodate the headwind, it diminishes and is replaced by a downdraft that is soon followed by a tailwind (and perhaps another ring vortex encounter). The rapid loss of airspeed and the low energy state of the aircraft both contribute to a trajectory undershoot that can lead to ground impact.

A good deal of research has been conducted on the wind shear hazard during the past decade. Much of it deals with the meteorological details of the phenomenon or general treatments of aircraft stability and *ad hoc* piloting effects [3-14]. Linear-quadratic regulators with feedforward control that provided remarkably good results in reduced-order linear simulations of wind shear encounter were developed [15]. This work illustrated the potential value of using "energy state", $E = V^2/2g + h$, either as a feedback variable or as an element of the quadratic cost function.

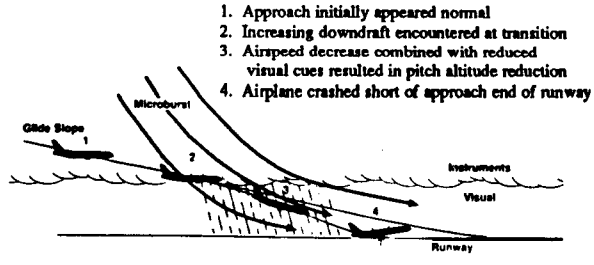


Figure 1. Wind shear encounter during approach (from "Windshear Training Aid," Federal Aviation Administration, Washington, DC, Feb 1987).

Additional applications of the energy state in aircraft guidance and control have been considered [16-22], and a Wind Shear Hazard Index that derives from the combined effects of the vertical shear in headwind and the downdraft magnitude on dE/dt was presented in Ref. 8. Following an extensive classical-control analysis, a feedback controller that dramatically reduces wind shear response by feeding flight path angle and the product of dynamic pressure and angle-of-attack perturbation to elevator and E and dE/dt to throttle was derived

[1] Professor.
[2] Research Assistant.

[23]. This study included realistic delays associated with short period, throttle, and elevator dynamics, which are important in real-world implementation.

Optimal flight paths and guidance strategies also have been suggested [24-29]. Knowing the optimal flight profiles corresponding to various wind fields is vitally important to designing and evaluating on-board systems to protect against wind shear, as survival may well depend on the specific control profile chosen by the pilot. Sub-optimal systems can provide an increased safety margin over conventional cockpit instrumentation, and a number of wind-shear-warning systems are commercially available. Nevertheless, one would like much more than sub-optimal protection through systems that properly anticipate dangerous conditions before they occur and that provide near-optimal guidance in inadvertent encounters. It is in this regard that concepts of machine intelligence may prove useful.

Background on Intelligent Guidance and Control

The human pilot is a most capable and robust controller, but the complexity, scope, and flight-criticality of making proper decisions and controlling the aircraft during wind shear encounter is an extreme challenge. The nature of the task is not, in any sense, beyond human understanding, but the need for assimilating a multitude of diverse data and taking correct actions may be beyond unaided human capability. Given time and information, an expert crew could arrive at a satisfactory solution; however, it is clear that computers could assess the quantitative aspects of the situation far more quickly than a human can. It would be desirable for a computer to evaluate symbolic, qualitative information as well, although the best way of doing this remains to be found.

A dynamic expert system will form a critical part of an on-board system to counter wind shear effects. Whereas static expert systems may consider fixed vectors of information, many components of the dynamic expert system's information vector are essentially time-varying, increasing the dimensionality of the information. As an example, consider the pattern-recognition problem. In the static case, a fairly straightforward discriminant function could be used to classify information; while in the dynamic case, a discriminant function must be applied to time series of data. In this instance, the typical "IF...THEN...ELSE" paradigm (in its simplest form) is inadequate, and hypothesis testing must be based on dynamic algorithms.

There are two principal approaches to interconnecting machine intelligence and automatic control, based on federated systems and integrated systems. In federated systems, the goal is to provide a symbolic supervisory control for a large-scale process control system, as used, for example, in the chemical industry. A LISP processor performing the first class of tasks could be connected to a conventional computer performing the second set of tasks. This decoupling of fundamentally different types of tasks provides good isolation for software development, and it facilitates the manual intervention of a skilled operator, should that be required. However, it is most appropriate for intrinsically slow operations. Integrated systems are Rule-Based Controllers (RBC) that combine symbolic and numeric processing in a highly integrated, parallel environment [30]. The principal characteristics of an integrated system are that numeric processes occur as side effects of forward and/or backward chaining, and the entire system is implemented for real-time execution in an array of similar processors coded with a procedural language.

Three areas appear to be critical to successful real-time implementation of machine intelligence in wind shear encounter. The first is acquisition of sufficient data on which to base a decision and to control the aircraft. The second is the definition of efficient hardware and software structures for a real-time expert system that addresses the wind shear problem. The third is the design of interfaces between the equipment and the pilot that are both complete and reasonable.

A real-time expert system for fault-tolerant control that has many of the features needed for an intelligent anti-windshear system has been developed [30-32]. That system provides a number of important pathfinders for such a system, including implementation of a controller with hierarchical structure, multiple cooperating expert systems in parallel microprocessors, and closure of an intelligent feedback control loop at a high sampling rate. The system structure has been developed using LISP, then translated semi-automatically to Pascal code for real-time operation on three 80286/7 Multibus boards. There are equivalent expert systems in the two languages, both of which access the same database and external simulation.

FAA WINDSHEAR TRAINING AID

The FAA Windshear Training Aid was prepared with the support of the Integrated FAA Wind Shear Program. This two-volume manual was written by a team from the airframe industry that interacted with airlines, government, and academia [33]. Principal results are expressed in a variety of ways for executive review, training classes, and public information. One principal goal is to identify the logical connections between pilot observations and pilot actions when wind shear is encountered. The functions that a jet transport aircraft crew should perform are summarized by a flow chart (Fig. 2).

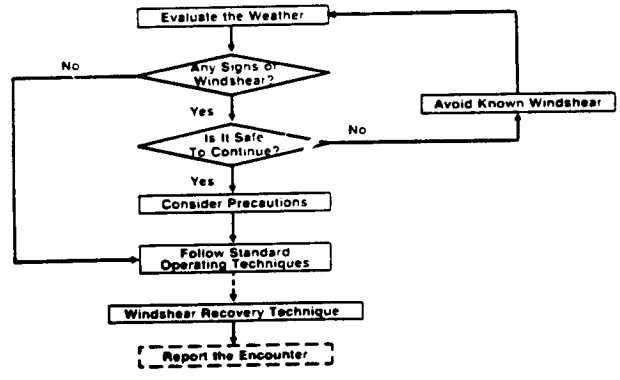


Figure 2. Model of flight crew actions. (from [33])

Flight crews are provided with guidelines that relate observations to wind shear probability along with examples of their use. If the probability of wind shear is LOW, standard procedures are recommended. If the probability is MEDIUM, the crew is instructed to consider precautions, including delay or alteration of terminal operations. If the probability is HIGH, delay or alteration of terminal operations is recommended, with specific actions guided by flight phase. If more than one observation suggests dangerous wind shear, the subjective probabilities should be added, although the guidelines for the risk assessment and the probability addition are imprecise. For example, two LOWs equal a MEDIUM, and either two MEDIUMs or a LOW and a MEDIUM equal a HIGH. There is no guidance regarding spatial or temporal characteristics of the observations; issues of proximity and degree of intensity are left to the pilot's judgment.

Although the strongest suggestion for piloting strategy is "avoid, avoid, avoid," recommended procedures for recovery or abort following wind shear encounter are given as

functions of flight phase. These strategies are sub-optimal, but they materially enhance the probability of survival, in comparison to standard piloting procedures.

The FAA Windshear Training Aid is a significant achievement in the fight against the hazards of low-altitude wind shear. Nevertheless, it takes a high level of piloting awareness and skill to evaluate the situation and to execute the implied actions correctly and quickly enough to avert catastrophe. To the extent that a computer can be fast and precise, it could assist the flight crew in this dangerous situation.

In seeking to build a computer aid for wind shear avoidance, it is necessary to model the implied logical patterns that the flight crew must use and to quantify subjective rules for computation. Many factors related to situational awareness, limitations to effective action, and efficient decision analysis must be considered, for the computer cannot exert "sound judgment," as suggested in the Training Aid, without having been programmed to do so.

WINDSHEAR SAFETY ADVISOR

The WindShear Safety Advisor (WSA) is a computer program that uses concepts drawn from the world of artificial intelligence (AI) to assess the wind shear threat and to recommend safe piloting action (Fig. 3). The current version is an

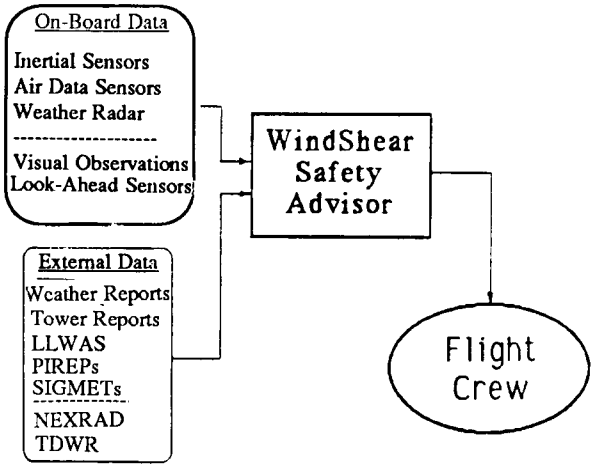


Figure 3. Structure of WindShear Safety Advisor.

interactive but non-real-time program for studying the input information and logic required to emulate and extend the FAA Windshear Training Aid to on-board computer systems. In particular, the WSA implements the stated rules of the Training Aid, and its development is uncovering the unstated (but critical) implications of the manual. The WSA, currently on a Symbolics 3670 LISP machine using the Genera environment, does not address important human factors issues, such as presentation of information to the pilot and requests for pilot input or intervention, which would have little significance in non-real-time simulation. However, our goal is to identify a program structure that is appropriate for real-time use.

As shown in Fig. 3, the WSA accepts data from sources that are external and internal to the aircraft. External sources include human controllers in an airport control tower, weather reporting systems, terminal doppler weather radar (TDWR), the low-level windshear alert system (LLWAS), and pilot reports (PIREPs). Internal data are transmitted from motion sensors, weather radar, visual observations, and (in the future) "look-ahead" sensors such as infrared, radar, or laser devices.

Functions of the WindShear Safety Advisor

The WindShear Safety Advisor, in its current state, is principally a model of pilot and crew decision-making and control, as described in the FAA Windshear Training Aid. The four primary functions of the system are:

MONITORING -	Observe sensors, receive reports and flight-plan revisions, alerts, and warnings
ASSESSMENT -	Detect wind shear encounters, determine if there are signs of wind shear, and if it is safe to continue
PLANNING -	Recommend actions and precautions to be taken
ACTION -	In automatic mode, execute standard, recovery, and go-around procedures; in semi-automatic mode, issue commands to flight directors

These functions are performed as *side effects* of a goal-directed search for parameter values in a set of rules.

An on-board implementation of the WSA would be a Rule-Based Control (RBC) system having attributes of both expert systems and conventional controllers (Fig. 4). A decision-making process implemented as an expert system may require that certain side tasks be accomplished, such as taking

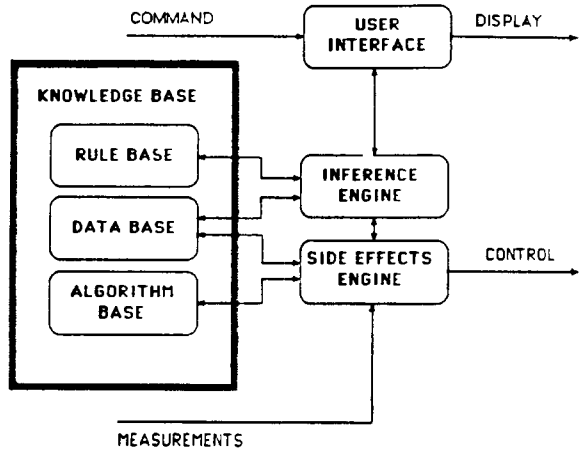


Figure 4. Rule-based system for control.

measurements, making estimates, computing control settings, and transferring commands to control effecters. This procedural, quantitative computation is done in a Side Effects Engine that calls on both the Data Base and an Algorithm Base for its knowledge. (Measurement and control are considered side effects of the request for information and the decision-making process.)

Decision and control functions are readily separated in an RBC system, the former calling for symbolic computation, the latter for numeric computation. (In either case, the digital computer simply moves bits around; however, interpretations of the logical operations are different.) Not surprisingly, some computer programming languages are better than others at performing the two types of tasks, so it is most efficient to use different languages for decision and control during the development phase. For example, LISP is a good language for developing logical relationships among strings of symbolic data, while Pascal or FORTRAN is a good language for numerical

computation. Consequently, LISP is the language of choice for current WSA development. Once decision and control functions have been defined, they must be merged (in some sense) in the RBC system. Development of a real-time version of an RBC system is thus aided by one or more language translators that efficiently transform subsets of the development languages into the final code.

Expert System Techniques and Implementation

An important feature of the RBC system is its use of a declarative representation of knowledge in a knowledge base. The rules and given facts are structured using *frames*. The elements of decision making needed for the WSA are illustrated by this simple example (Fig. 5). A *parameter* is a quantity that

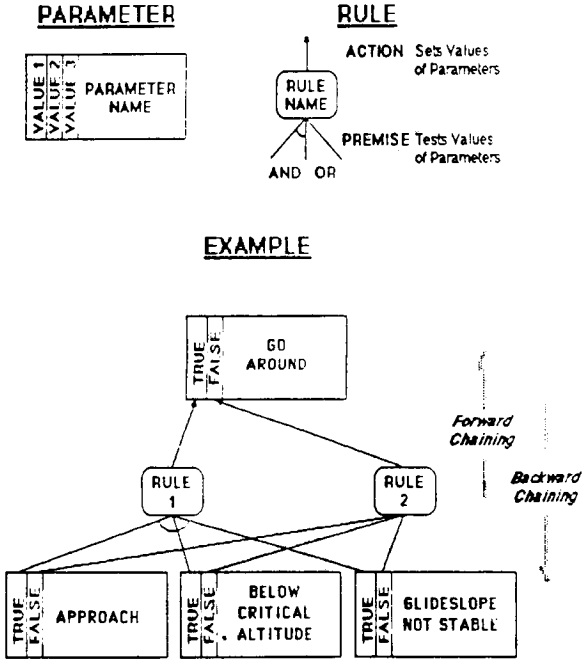


Figure 5. Graphical representation of knowledge.

can have several values as well as an array of complicated properties (not shown). A *rule* accepts one or more parameters as its *premise* and performs the *action* of setting another parameter if the values of its input parameters make the rule true. For the premise to be true, it may be necessary that *all* multiple parameters take certain values (represented by the arc between connecting lines), or it may be sufficient for *any* parameter to take a certain value (represented by no arc between channels into the rule).

The process of applying the knowledge base to given facts is known as *inference*. Given a base set of facts and rules, all of the facts that follow may be asserted by examining all the rules, in any order, repeatedly, until no new facts can be proven. However, the order in which the rules are tried can make a big difference in the efficiency of this process. With a large number of rules (which is always the case for any substantial amount of expert knowledge), efficient inference in an expert system requires a strategy for deciding the order in which the rules are examined.

Two different types of inference procedures are used in the WindShear Safety Advisor. These procedures can be distinguished by how the expert system obtains facts from the user. With one type of procedure, the user gives a set of facts to the system, and the procedure works forward from these facts, deriving all the possible consequences of them. This type of procedure is called *data-driven* or *forward-chaining inference*. With the other type of procedure, the user directs the system to find the truth or falsehood of a goal parameter, and the search procedure works backward from this goal parameter, asking the user to supply values for parameters whose values are not known, relevant to the goal parameter, and cannot be given values by examining rules. This type of procedure is often called *goal-directed* or *backward-chaining inference*. The WindShear Safety Advisor uses both types of procedures. Taken together, these procedures are referred to as the *inference engine*.

To demonstrate the actions of goal-directed inference, consider the following example (Fig. 6). The expert system is

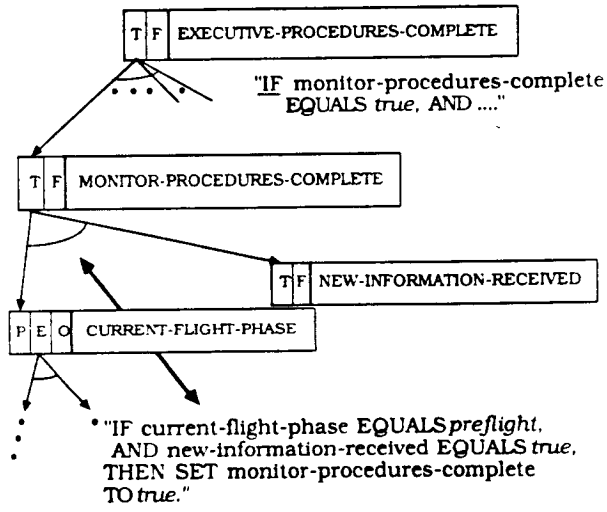


Figure 6. Graphical representation of goal-directed inference.

invoked periodically from a simulation by directing the inference procedure to get a value for the parameter "Executive-procedures-complete". "Executive-procedures-complete" is set in the EXECUTIVE rule base by the top-most rule in the figure, which considers the truth or falsehood of the expression, "Monitor-procedures-complete". To know this, we must find out what the flight phase is. Turning to the list of parameters, we find a parameter "Current-flight-phase" that is set in the FLIGHT PHASE rule base. The rules within the FLIGHT PHASE rule base are examined next. When the flight phase has been determined, we return to considering the truth or falsehood of the first premise. If, for example, we determined that the flight phase was "Preflight," we would then consider whether or not new information had been received. Once this search is complete, the parameter "Executive-procedures-complete" is set to TRUE, and the top-level search is finished.

A data *frame* is a structure that defines the attributes of objects, as well as their linkages to other frames in the knowledge base. We might, for example, describe a hazardous volume of airspace as a sphere of given location and radius (Fig. 7). The *sphere frame* has two slots: location of the center and radius. The latter is specified by a number, while the former is itself another frame that contains four slots, and so on. Each sphere is denoted by a frame *instance*, in which actual values appear in the slots. When a specific sphere is defined, it is said to be "instantiated."

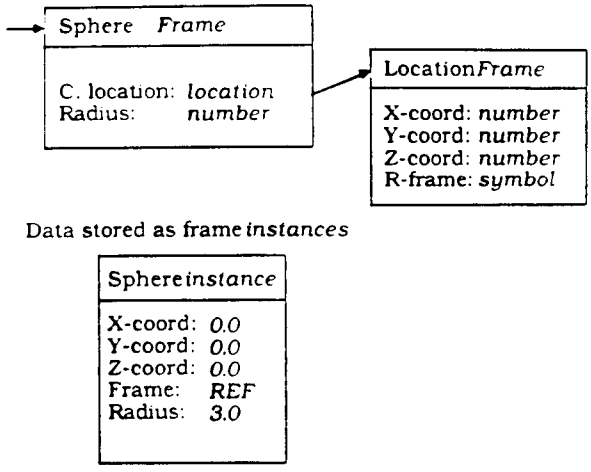


Figure 7. Structuring data with frames.

The WSA contains numerous other frames that describe the aircraft's dynamic state and its so-called "world environment." For example, the flight phase, event, and hazard frames are defined as follows, where terms in italics refer to specific data types:

Flight Phase *Frame*

Type:	*f-type*
Airport:	*airport*
Runways:	*rnwy-list*
Time interval:	*tinterval*
Nominal trajectory:	*tp-list*

Event *Frame*

Type:	*event-type*
Reported by:	*info-source*
Intensity:	*int-metric*
Space-time description:	*std-type*

Hazard *Frame*

Type:	*h-type*
Sphere:	*std-type*
Affected trajectory:	*t-points*
Risk factor:	*number*

Qualitative decisions must be made on the basis of information contained in these frames, the risk of wind shear encounter providing a good example. An appropriate rule is

IF there is an instance of event type microburst,
AND we intersect it in the next flight phase,
AND a value for severity is obtained from the microburst-indicator rule base,

THEN a microburst hazard is instantiated,
AND the risk factor is increased by the incremental severity value.

The Windshear Training Aid guidelines provide *ad hoc* ways of determining incremental severity values, but a great deal is left to the implied good judgment of the human decision-maker. Computer-based decisions require an approach that is at least quasi-quantitative if not rigorously so. For example, the LOW-MEDIUM-HIGH descriptors of wind shear probability can be equated to an ordinal ("1-2-3") scale to implement the above logic. As increasingly numeric information is gathered about risk factors, fractional severity metrics then can be incorporated. Rainshowers on the flight path might reasonably indicate a higher concern than showers near the path but not high enough to increment the risk factor by an integer value.

Another important feature of any expert system is an interface to translate a user's problem into a collection of symbols and to translate the results of inference back into a format the user can understand. The WindShear Safety Advisor Interface, used for logic development and demonstration, is a multiple-window interactive display (Fig. 8). The current version accepts keyboard and mouse input, with the results of inference displayed as text output. Execution of the system consists of an elementary dialogue between the program user and the expert system. As the user answers the system's questions, statements of applicable rules are typed out in one window, while information about parameters is displayed in another. Special messages summarizing the system's advice are printed in a third window.

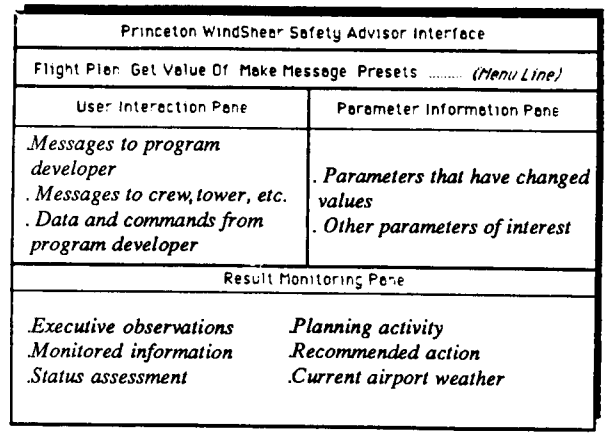


Figure 8. Development screen for WindShear Safety Advisor.

CONCLUSIONS

The WindShear Safety Advisor program implements the stated decision-making logic of the FAA Windshear Training Aid, as well as a set of unstated implications that are necessary for practical application. The WSA expert system contains over 200 rules that set over 80 parameters for terminal operations of jet transport aircraft. Future modifications will account for spatial and temporal variations of the aircraft and its meteorological environment, as well as for interfaces with the air traffic control system. The WindShear Safety Advisor sets the stage for cockpit simulation of logic for wind shear avoidance, which, in turn, will lead to practical systems for operational aircraft.

ACKNOWLEDGMENT

This research has been sponsored by the NASA Langley Research Center under Grant No. NAG-1-834.

REFERENCES

1. Townsend, J., committee chairman, "Low-Altitude Wind Shear and Its Hazard to Aviation," National Academy Press, Washington, 1983.

2. Stengel, R., "Solving the Pilot's Wind-Shear Problem," *Aerospace America*, Vol. 23, No. 3, Mar 1985, pp. 82-85.

3. Fujita, T., "Downbursts and Microbursts - An Aviation Hazard," *Proceedings of the 19th Conference on Radar Meteorology*, Miami Beach, 1980, pp. 94-101.

4. McCarthy, J., *et al*, "The Joint Airport Weather Studies (JAWS) Project," *Bulletin of the American Meteorological Society*, Jan 1982, pp. 15-22.

5. Bray, R., "A Method for Three-Dimensional Modeling of Wind-Shear Environments for Flight Simulator Applications," NASA TM-85969, July 1984.

6. Zhu, S., and Etkin, B., "Fluid-Dynamic Model of a Downburst," UTIAS Report No. 271, Apr 1983.

7. Bach, R., Jr., and Wingrove, R., "The Analysis of Airline Flight Records for Winds and Performance with Application to the Delta 191 Accident," AIAA Paper No. 86-2227-CP, Aug 1986.

8. Bowles, R., "Fundamental Concepts for Wind Shear Detection, Warning and Flight Guidance," presentation at Airborne Wind Shear Systems Manufacturers Review," Hampton, Oct 1986.

9. Sherman, W., "A Theoretical Analysis of Airplane Longitudinal Stability and Control as Affected by Wind Shear," NASA TN D-8496, July 1977.

10. Gera, J., "The Influence of Vertical Wind Gradients on the Longitudinal Motions of Airplanes," NASA TN D-6430, Sept 1971.

11. Frost, W., "Flight in Low-Level Wind Shear," NASA CR-3678, Mar 1983.

12. Lehman, J., *et al*, "Simulation and Analysis of Wind Shear Hazard," FAA-RD-78-7, Dec 1977.

13. Urnes, J., *et al*, "H-Dot Automatic Carrier Landing System for Approach Control in Turbulence," AIAA *Journal of Guidance and Control*, Mar-Apr 1981, pp. 177-183.

14. Foy, W., "Airborne Aids for Coping with Low-Level Wind Shear," FAA-RD-117, July 1979.

15. Rynaski, E., and Govindaraj, K., "Control Concepts for the Alleviation of Windshears and Gusts," NASA CR-166022, July 1982.

16. König, R., *et al*, "Procedures to Improve Flight Safety in Wind Shear Conditions," *Proceedings of the 12th Congress of the International Council of the Aeronautical Sciences*, ICAS-80-22.3, Munich, Oct 1980, pp. 744-757.

17. Lambregts, A., "Vertical Flight Path and Speed Control Autopilot Design Using Total Energy Principles," AIAA Paper No. 83-2239-CP, Aug 1983.

18. Belcastro, C., and Ostroff, A., "Total-Energy-Rate Feedback for Automatic Glide-Slope Tracking During Wind-Shear Penetration," NASA TP-2412, May 1985.

19. Ostroff, A., *et al*, "Evaluation of a Total-Energy-Rate Sensor on a Transport Airplane," NASA TP-2212, Nov 1983.

20. Jorck, H., "Design of Wind Shear Filters," NASA TM-77460, Nov 1984.

21.Broussard, J., "Extensions to PIFCGT: Multirate Output Feedback and Optimal Disturbance Suppression," NASA CR-3968, Mar 1986.

22. Hahn, K., "Take-Off and Landing in a Downburst", *Proceedings of the 15th Congress of the International Council of the Aeronautical Sciences*, ICAS-86-5.6.2, Sept 1986, pp. 860-869.

23. Psiaki, M., and Stengel, R., "Analysis of Aircraft Control Strategies for Microburst Encounter," AIAA *Journal of Guidance, Control, and Dynamics*, Vol. 8, No. 5, Sept-Oct 1985, pp. 553-559.

24. Miele, A., Wang. T., and Melvin, W., "Optimal Take-Off Trajectories in the Presence of Windshear," *Journal of Optimization Theory and Application*, Apr 1986, pp. 1-45.

25. Psiaki, M., and Stengel, R., "Optimal Flight Paths Through Microburst Wind Profiles," AIAA *Journal of Aircraft*, Aug 1986, pp. 629-635.

26. Psiaki, M., "Control of Flight Through Microburst Wind Shear Using Deterministic Trajectory Optimization," Ph.D. Thesis, Princeton University, 1987.

27. Miele, A., Wang, T., and Melvin, W., "Optimization and Gamma/Theta Guidance of Flight Trajectories in a Windshear," *Proceedings of the 15th Congress of the International Council of the Aeronautical Sciences*, ICAS-86-5.6.4, Sept 1986, pp. 878-899.

28. Stengel, R., "Optimal Control Laws for Microburst Encounter," *Proceedings of the 15th Congress of the International Council of the Aeronautical Sciences*, ICAS-86-5.6.3, Sept 1986, pp. 870-877.

29. Psiaki, M., and Stengel, R. "Performance Limits for Optimal Microburst Encounter, *Proceedings of the 1988 AIAA Atmospheric Flight Mechanics Conference*, Minneapolis, Aug 1988, pp. 358-370.

30. Handelman, D., and Stengel, R., "An Architecture for Rule-Based Control," *Proceedings of the 1987 American Control Conference*, Minneapolis, June 1987, pp. 1636-1642.

31. Handelman, D., and Stengel, R., "Combining Quantitative and Qualitative Reasoning in Aircraft Failure Diagnosis," *Proceedings of the 1985 AIAA Guidance, Navigation, and Control Conference*, AIAA Paper No. 85-1905CP, Aug 1985.

32. Handelman, D., and Stengel, R., "A Theory for Fault-Tolerant Control Combining Expert System and Analytical Redundancy Concepts, *Proceedings of the 1986 AIAA Guidance, Navigation, and Control Conference*, AIAA Paper No. 86-2092CP, Aug 1986.

33. -- , "Windshear Training Aid," Federal Aviation Administration, Washington, DC, Feb 1987.

RULE-BASED MECHANISMS OF LEARNING FOR INTELLIGENT ADAPTIVE FLIGHT CONTROL

David A. Handelman* and Robert F. Stengel**

Princeton University
Department of Mechanical & Aerospace Engineering
Princeton, New Jersey

ABSTRACT

This paper investigates how certain aspects of human learning can be used to characterize learning in intelligent adaptive control systems. Reflexive and declarative memory and learning are described. It is shown that model-based systems-theoretic adaptive control methods exhibit attributes of reflexive learning, whereas the problem-solving capabilities of knowledge-based systems of artificial intelligence are naturally suited for implementing declarative learning. Issues related to learning in knowledge-based control systems are addressed, with particular attention given to rule-based systems. A mechanism for real-time rule-based knowledge acquisition is suggested, and utilization of this mechanism within the context of failure diagnosis for fault-tolerant flight control is demonstrated.

INTRODUCTION

Adaptability is an essential feature of any control system designed to interact effectively with the real world. Uncertainty motivates much of the need for adaptability, directly affecting control system stability and performance. Sources of uncertainty are many, representing inadequacies in knowledge of the system to be controlled, or in the environment within which the system must operate. For example, uncertainty can result from corruption of incoming information due to sensor noise or failure. It can also result from changes in control effectiveness due to failure or unanticipated changes in the operating environment, changes in system dynamics due to environmental factors or structural failure, and unanticipated or poorly-modeled external disturbances. Control law design must address to some degree these issues.

Fortunately, effective control techniques capable of accommodating certain types of uncertainty exist. Robust stochastic optimal estimation and control methods, for example, perform well in the presence of Gaussian sensor noise and state disturbances [1]. Characteristic changes in the dynamics of the controlled system often can be accommodated using parameter estimation and adaptive control techniques [2]. Under certain circumstances, self-organizing controllers may be used to perform non-trivial tasks given little prior information about the kinematics and dynamics of the system being controlled [3,4,5,6].

* Graduate Student
** Professor
Presented at the 1988 American Control Conference, Atlanta, June 1988.

Even the most accommodating numerical control techniques, however, be they robust, adaptive, or self-organizing, eventually reach limits of performance when deficiencies in knowledge of the plant or its environment exceed certain thresholds. One limiting factor relates to the ability of the control system to learn about novel, important relationships and events in the world, and how to respond properly to them. Machine learning is an active area of research in the field of Artificial Intelligence (AI) [7,8,9,10,11]. However, although problem-solving techniques of AI are finding their way into various phases of control system design and implementation [12,13], little work has addressed the issue of learning in demanding real-time applications such as aircraft and spacecraft flight control.

This paper investigates how certain aspects of human learning can be used to characterize learning in "intelligent" control systems. Two types of memory and learning are described. It is shown that model-based adaptive control methods are particularly well suited for implementing one type, whereas the problem-solving capabilities exhibited by knowledge-based systems of artificial intelligence make them naturally suited for implementing the other type. Issues related to learning in knowledge-based control systems are addressed, with particular attention given to rule-based systems. A mechanism for rule-based knowledge acquisition is suggested, and utilization of this mechanism within the context of failure diagnosis for fault-tolerant flight control is described.

REFLEXIVE AND DECLARATIVE MEMORY AND LEARNING

Learning relates to knowledge acquisition, memory to its storage. Various methods of classification are used by psychologists to describe different types of memory and learning exhibited by humans. One classification scheme is based on how learned information is encoded and recalled, distinguishing between what some authors term reflexive and declarative memory and learning [14]. With regard to control, maneuvers indicative of reflexive mechanisms may be characterized as automatic, requiring little or no thought. Maneuvers involving declarative memory and learning, on the other hand, require conscious effort. Evaluation, comparison, and inference characterize declarative thinking. Moreover, whereas reflexive learning relates specific responses to specific stimuli, declarative learning provides insight into not only how something is done, but why. Any complex task attempted for the first time involves some form of declarative reasoning.

Reflexive and declarative memory and learning are closely related. Tasks initially learned declaratively often become reflexive through repetition. Conversely, when familiar tasks are attempted in novel situations, reflexive knowledge must be converted back into declarative form in order to become useful. For example, although one may become adroit at tying one's own necktie, tying someone else's necktie requires some thought due the change in perspective. By drawing analogies to such human information processing mechanisms, adaptive control systems might benefit from the incorporation and integration of both reflexive and declarative forms of learning [15].

A simplistic example of learning pertinent to aircraft flight control demonstrates that distinctions between reflexive and declarative learning can affect aircraft stability and performance. Consider what happens when a student pilot is taught how to recognize and recover from a wings-level approach-to-landing stall. The student is shown how sluggish control response and aircraft buffeting indicate low airspeed and impending stall, and that recovery includes pushing the control stick forward. Reflexive learning would encode knowledge similar to the following.

```
If    control response is sluggish
      and
      buffeting is encountered
then  push stick forward
```

Conversely, declarative learning would result in the acquisition of knowledge encoding more causal detail, such as the following.

```
If    control response is sluggish
      and
      buffeting is encountered
then  decrease magnitude of angle of attack

If    magnitude decrease in angle of attack
      is required
then  push stick forward
```

The type of knowledge acquired by the student, reflexive or declarative, has a large impact on the student's ability to apply this knowledge in novel situations (and hence to adapt). For example, during aerobatic flight, the student notices significant differences between required control inputs for inverted and non-inverted flight.

```
If    aircraft is upright
      and
      magnitude decrease in angle of attack
      is required
then  push stick forward

If    aircraft is inverted
      and
      magnitude decrease in angle of attack
      is required
then  pull stick aft
```

How will the student respond when pre-stall conditions (control sluggishness and airframe buffeting) are encounted during inverted flight? A reflexive response would be based on the relationship between pre-stall warnings and forward stick movement learned during non-inverted flight. Such a response would aggravate the stall. Declarative thinking, on the other hand, would recognize the need for a reduction in the magnitude of angle of attack, and that when inverted such a reduction is accomplished with aft stick movement.

The distinction between reflexive and declarative memory and learning suggests what roles existing systems-theoretic adaptive control techniques and proposed artificial intelligence methodologies can play in adaptive control systems. Most existing adaptive control techniques are based (for good reason) on mathematical models of the dynamic system being controlled. The analytical functions representing the adaptive control law ultimately calculate specific control commands in response to specific sensor measurements. In this sense, these model-based techniques can be viewed as implementing reflexive knowledge. Conversely, the inferencing capability of knowledge-based systems can be viewed as implementing declarative knowledge. Many model-based adaptive control techniques benefit from the availability of closed-form analytical expressions dictating how model parameters should be modified. Learning often proceeds in a stable and optimal fashion, and these algorithms should be used when possible. Unfortunately, with knowledge-based systems, no such rigorous guidelines for learning exist.

In general, a learning controller should be able to identify important new information, decide whether this new information should augment or replace existing knowledge, and transfer this information into the existing knowledge base. Possible schemes for knowledge acquisition include rote learning, learning by example, and learning by trial and error. As mentioned above, much AI research addresses learning in knowledge-based systems, and the field of adaptive control most likely will benefit from its advances. Any knowledge-based application, however, is based upon a specific form of knowledge representation. The following sections suggest that rule-based expert system techniques provide a sound representational basis for declarative learning in time-critical adaptive control systems.

REAL-TIME DECLARATIVE LEARNING THROUGH RULE RECRUITMENT

The knowledge representation and problem-solving features of rule-based systems make them particularly well-suited for implementing the causal relationships characteristic of declarative learning. Moreover, as will be demonstrated, knowledge acquisition can be made to occur in the computationally efficient manner required for real-time control. Note that the discussion below focuses on mechanisms enabling automated rule-based knowledge acquisition, not on how this new knowledge is identified.

Within a certain class of forward- and backward-chaining rule-based systems [16], the knowledge base is composed of parameters and rules. Parameters represent symbolic information. Each parameter may acquire one of a list of allowed values, or its value may be considered unknown. Information expressing relationships and dependencies between parameter values is contained in rules. Each rule contains a premise and an action. If a rule premise is true

when tested, its action is executed, causing the inference of additional information. In control system applications, rule actions also may perform specific control tasks. Rule testing is guided by an inference engine, an operator applied to the knowledge base enabling the process of search. Parameters thereby represent a partial description of the "state of the world", and rule-based search is used to modify this description. Figure 1 defines a symbology useful in graphically depicting such a knowledge base. Rectangles contain all values that the corresponding parameter can acquire. With arcs between parameters representing rules, the resultant AND/OR graph can be used to trace the logic path taken by the search process.

Various forms of search may be applied to the knowledge base. The key to search is the manner in which rules are linked through parameters. Lists of rule names associated with each parameter provide this link. For example, the purpose of a goal-directed (backward-chaining) search is to infer a value for a specified parameter. To this end, each parameter has attached to it a list identifying which rules are capable of modifying the value of the parameter through the rule action. This list is consulted by the inference engine during goal-directed search when a parameter value must be inferred.

In general, knowledge acquisition within such a rule-based system involves three steps. These three steps are depicted graphically in Fig. 2 to 4 with reference to the knowledge associated with the stall recovery scenario given above. First, parameters capable of representing the "state space" of knowledge to be learned are collected as shown in Fig. 2. In the second knowledge acquisition step, rules associating parameter states are constructed as in Fig. 3. Finally, rules are linked by updating the rule-name lists associated with relevant parameters. The resultant knowledge base is depicted in Fig. 4.

Following incorporation into the knowledge base, these new rules may be utilized for control system problem solving. A goal-directed search on the knowledge of Fig. 4, for example, would begin with the question, "How should the stick be moved?" or more specifically the instruction, "Determine the value of parameter DESIRED_STICK_MOVEMENT." Rules 1 and 2 are capable of supplying this information. Rule 1 is tested first, with its premise initially checking aircraft attitude. Assume the aircraft is inverted. In this case, Rule 1 fails, and Rule 2 is tested. The premise of Rule 2 eventually needs to know whether or not a magnitude decrease in angle of attack is required. Rule 3 is capable of supplying this information; therefore it is tested at this time. If control response is sluggish and buffeting is encountered, the action of Rule 3 determines that a change in angle of attack is required, finally allowing the action of Rule 2 to determine that aft stick movement is appropriate.

Rules represent executable code. Knowledge acquisition as specified above involves the automatic generation and execution of code during control system operation. List-based programming languages such as LISP can be used to accomplish such feats. Problems can arise, however, when symbolic programming languages and hardware must be integrated with

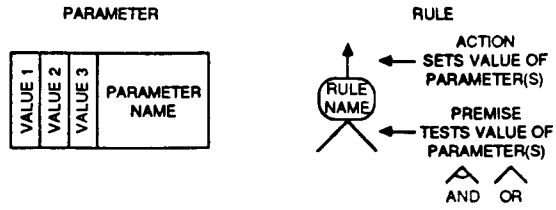


Fig. 1. Graphical Representation of Rule-Based Knowledge.

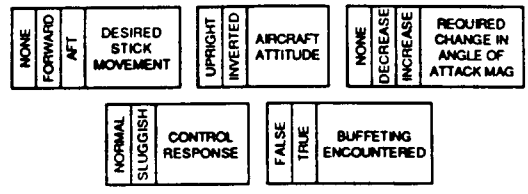


Fig. 2. Parameter Representation of Acquired Sample Knowledge.

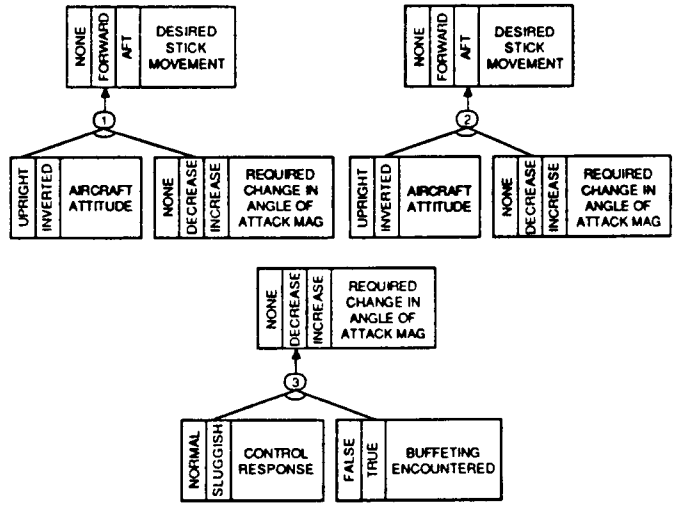


Fig. 3. Rule Representation of Acquired Sample Knowledge.

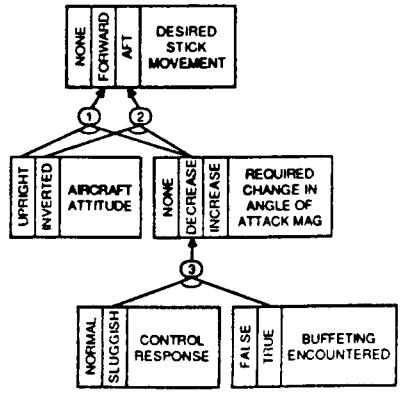


Fig. 4. Knowledge Base Representation of Acquired Sample Knowledge.

numeric control system processing chores [17]. Fortunately, by sacrificing some flexibility, real-time performance can be obtained. For example, by encoding rules in a structured procedural language such as C or Pascal, and integrating them with numerical routines in a multiprocessor system, symbolic processing capability remains limited, yet powerful [18,19]. Furthermore, knowledge acquisition can be enabled in such a system through a process called rule recruitment.

Rule recruitment involves the manipulation of a stack of dormant rules. A rule is defined as dormant if it cannot be referenced during search. This situation will occur if the rule's name does not appear on any parameter's rule-name list. For example, before the rules depicted in Fig. 3 are linked into the knowledge base through their respective parameters, they remain dormant and inaccessible to the inference engine. Simply by manipulating parameter lists, rules may be transferred into and out of dormancy.

Rules on the dormant rule stack serve as rule "templates". Each retains a fixed premise and action structure but refers to parameters and parameter values indirectly through pointer-type references. For example, the following rule template exhibits a structure capable of encoding the rules of Fig. 3.

```
If    <parameter pointer> is <value pointer>
      and
      <parameter pointer> is <value pointer>
then  set <parameter pointer>
      to  <value pointer>
```

By manipulating rule template pointers, rules may be built automatically as required. The process of rule recruitment, therefore, involves pulling a rule template off the dormant rule stack, initializing its pointers so that it encodes the desired chunk of knowledge, and modifying the appropriate parameter lists so that the new rule becomes an active part of the knowledge base. Figure 5 depicts the process of rule recruitment.

The major drawback with this knowledge acquisition mechanism is the inability to build arbitrarily complex rules at run-time. All desired parameters must be pre-defined as well. However, by forcing a system designer to formalize the structure of knowledge to be learned by the controller, these limitations may prove beneficial in the long run. A restricted set of unique rule templates encourages modular construction of more complex rules. Furthermore, a large set of unique rule templates may be designed into the dormant rule stack if needed. This remedy is memory intensive, not computation intensive, and memory is inexpensive.

The major advantage associated with rule recruitment is that it is fast. The execution time incurred during pointer assignment and parameter list updating is negligible. Moreover, the recruitment of rules can be overseen by other rules dedicated to knowledge acquisition, in much the same way that meta-rules can be used to guide rule selection during search [16]. Consequently, the knowledge acquisition mechanism of rule recruitment fits neatly into the existing computationally efficient rule-based control system environment.

APPLICATION OF RULE RECRUITMENT TO FLIGHT CONTROL SYSTEM DESIGN

A demanding adaptive control application was chosen as a testbed for some of the rule-based learning ideas presented above. The Rule-Based Flight Control System (RBFCS) is designed to combine analytical redundancy and expert system concepts for fault-tolerant flight control [19,20]. The software and hardware architectures of the RBFCS provide for real-time integrated symbolic and numeric processing. Within this setting, rule-based learning has been used in conjunction with model-based simulation to facilitate certain phases of control system design.

The RBFCS is intended to detect, identify, and reconfigure for a wide range of aircraft failures. The overall job of failure accommodation is broken down into five main tasks. The Executive Control Task provides continual dynamic state estimation, feedback control calculations, and synchronization of the remaining tasks. The Failure Detection Task monitors aircraft behavior and detects significant abnormalities. The Failure Diagnosis Task finds a set of probable causes of the problem, and the Failure Model Estimation Task generates a mathematical model of the aircraft dynamics considered to reflect changes arising from the assumed failure. Finally, the Reconfiguration Task determines what action should be taken to correct the situation. Automated learning has been used to generate rules relevant to failure diagnosis.

At the core of the Failure Model Estimation Task is a numerical algorithm that chooses from among a group of failure hypotheses the one most likely (in a probabilistic sense) to represent the actual failure. The number of hypotheses considered by the algorithm must be kept low, and this is the job of the Failure Diagnosis Task, which performs initial failure candidate screening. The intent is to use expert system techniques to emulate in real-time the reasoning of pilots, engineers, and mechanics familiar with the aircraft in order to make informed judgments as to what did or did not fail [21].

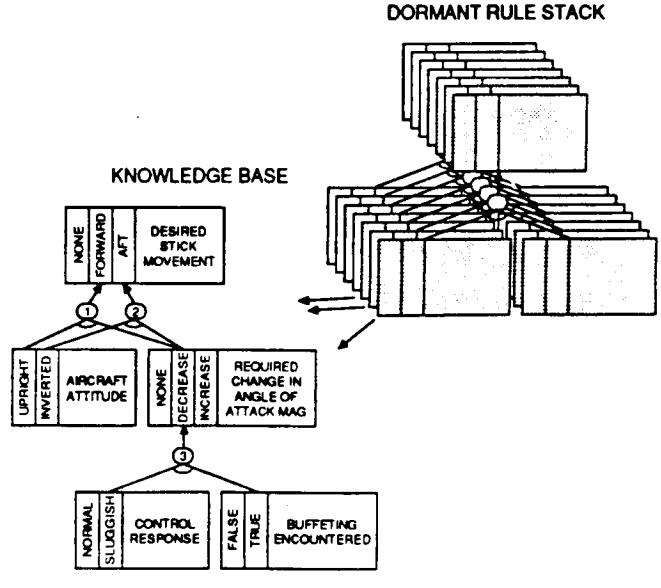


Fig. 5. Rule Recruitment Learning Mechanism.

In addition to containing explicitly specified heuristic knowledge, the Failure Diagnosis Task also has been given the capability of learning through simulation how the Failure Detection Task will respond to various failures. The impetus behind this capability is the intent to accommodate eventually an extremely large number of possible failure modes. The RBFCS presently is configured to accommodate a biased or stuck sensor or control in a U. S. Army CH-47 tandem-rotor helicopter travelling at 80 knots airspeed and sea level altitude. However, accommodation of structural failures affecting aircraft dynamics, as well as multiple and intermittent failures in sensors and controls, is included in the design goal of the RBFCS. By using analytical models of these failures, stochastic Monte Carlo type simulations can be used to characterize the effect that such complex failures have on failure detection. Learning by example off-line, the Failure Diagnosis Task generates a set of rules that approximates the effect of each failure mode. Utilizing these rules on-line, the task bases its initial screening of failure candidates, in part, on similarities between available failure-time information and effects known to be caused by specific types of failures.

This simulation-based learning by example is accomplished using rule recruitment. Presently, seven rules are used to approximate the average effect each failure mode has on 16 indicators. Windowed average and root-mean-square values of the residuals of an 8-state estimator are used as indicators. The 24 possible failure modes correspond to abrupt bias and stuck failures in 8 sensors and 4 controls. Sensors measure body-axis longitudinal velocity, lateral velocity, vertical velocity, roll rate, pitch rate, yaw rate, pitch angle, and roll angle. Controls include the two actuators of each rotor: forward cyclic pitch, forward collective pitch, aft cyclic pitch, and aft collective pitch. Each rule has the following form.

```
If    indicator_01 is near x.x
      and
      indicator_02 is near x.x
      and
                  :
      and
      indicator_16 is near x.x
then  there is good chance that
      forward collective pitch control is
      biased from nominal by an amount near x.x
      and
      failure detection delay is near x.x
```

The definition of "near x.x" in these rules is defined by fuzzy functions [22]. Failure detection delay is the time difference between detection and occurrence of the failure. Heuristics are used to distribute the 7 rules per failure mode throughout the expected failure mode range.

When recruited into the knowledge base, the 168 failure-effect rules integrate features of function approximation and pattern recognition with the remaining heuristics of the Failure Diagnosis Task. They help estimate at failure detection time the relative likelihood of each failure mode, failure mode magnitude and direction, and failure detection delay. The relative likelihood of a failure mode depends on the validity of its rule's premises, and it is used to narrow the initial set of failure mode candidates down to a reasonable size.

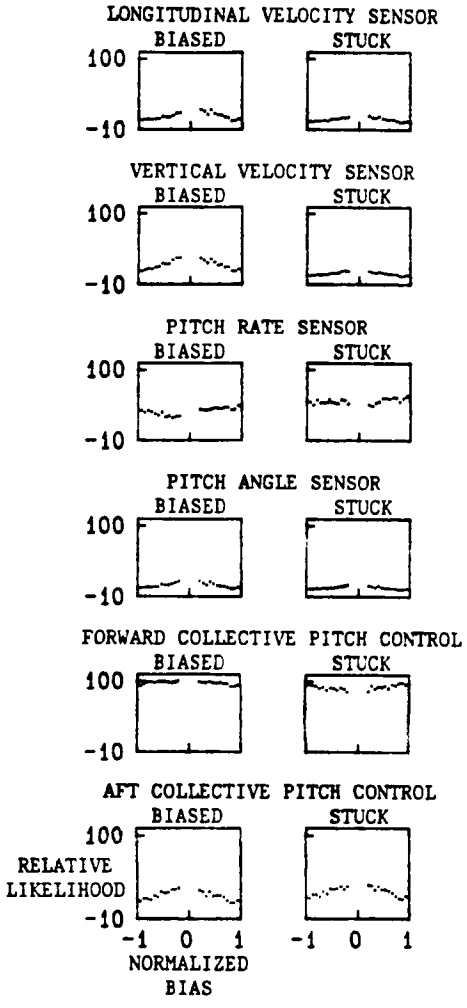


Fig. 6. Rule-Based Failure Diagnosis Performance: Failure Mode Likelihoods Given Biased Forward Collective Pitch Control.

Figure 6 illustrates some of the on-line information generated by these off-line recruited rules. For this figure, bias failures in forward collective pitch control were simulated using a stochastic fully-coupled 8th-order linear model of aircraft dynamics. Failure Diagnosis Task processing typically required less than 3 seconds on a computer equipped with an 8-MHz 80286 CPU and an 8-MHz 80287 math coprocessor. The abscissa of each plot in the figure corresponds to the amount of failure bias, normalized for a range of +/- 7.5 cm. Each data point in a plot corresponds to the average relative likelihood of the associated failure mode candidate over 20 simulated failure runs. Failure detection delays for these bias failures varied between 0.2 sec and 2 sec. Failures remaining undetected 2 sec after their occurrence were considered undetectable, explaining the gap in data associated with near-zero bias failures. It can be seen from the plots in Fig. 6 that likelihoods associated with the actual failure mode described as "forward collective pitch control biased" usually were the highest. Furthermore, other failure mode candidates capable of strongly affecting longitudinal dynamic state variables had significant likelihoods, as expected. Likelihoods associated with failure modes strongly affecting lateral/directional states (not shown) remained low.

By applying other heuristics in concert with this type of information, the Failure Diagnosis Task can quickly narrow the number of failure mode candidates from 24 down to 6 or less. Note that due to a rule-based implementation, the failure-effect knowledge generating this information can be used on a conditional basis if desired. For example, rule premises can be made sensitive to previously identified failures. The rule-based technique thereby exhibits in this case certain advantages over standard pattern matching techniques. Additionally, although the recruited failure diagnosis rules of the RBFCS were obtained off-line as part of control system design, the same mechanism of learning could be utilized during on-line control system operation.

CONCLUSIONS

The problem-solving capabilities of numeric model-based systems and symbolic knowledge-based systems can be used to implement various forms of automatic learning. The concept of learning through rule recruitment described above serves as an extension to work originally designed to integrate such symbolic and numeric processing for real-time control. Rule recruitment provides a mechanism whereby knowledge may be acquired automatically in a timely manner, allowing rules to generate additional rules. It can be used as a representational vehicle through which more fundamental issues of control system learning may be addressed, such as the acquisition and maintenance of general knowledge for highly adaptive aircraft and spacecraft flight control.

ACKNOWLEDGMENT

This project was sponsored by the U. S. Army Research Office under Contract No. DAAG29-84-K-0048.

REFERENCES

1. Stengel, R., Stochastic Optimal Control: Theory and Application, John Wiley & Sons, New York, 1986.

2. Goodwin, G., and Sin, K., Adaptive Filtering Prediction and Control, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1984.

3. Saridis, G., Self-Organizing Control of Stochastic Systems, Marcel Dekker, Inc., New York, 1977.

4. El-Fattah, Y., and Foulard, C., Learning Systems: Decision, Simulation, and Control, Springer-Verlag, New York, 1978.

5. Narendra, K., ed., Adaptive and Learning Systems: Theory and Applications, Plenum Press, New York, 1986.

6. Miller, W. III, "Sensor-Based Control of Robotic Manipulators Using a General Learning Algorithm," IEEE J. Robotics and Automation, Vol. RA-3, No. 2, April 1987.

7. Barr, A., Cohen, P., and Feigenbaum, E., The Handbook of Artificial Intelligence, William Kaufmann, Los Altos, California, 1982.

8. Michalski, R., Carbonell, J., and Mitchell, T., Machine Learning: An Artificial Intelligence Approach, Tioga Publishing Company, Palo Alto, 1983.

9. Michalski, R., Carbonell, J., and Mitchell, T., Machine Learning: An Artificial Intelligence Approach Volume 2, Morgan Kaufmann Publishers Inc., Los Altos, California, 1986.

10. Michalski, R., Carbonell, J., and Mitchell, T., Machine Learning: A Guide to Current Research, Kluwer Academic Publishers Inc., Boston, 1986.

11. Widrow, B., Winter, R., and Baxter, R., "Learning Phenomena in Layered Neural Networks," Proc. First Internat. Conf. on Neural Networks, San Diego, June 1987.

12. Hankins, W., Pennington, J., and Barker, L., "Decision-Making and Problem-Solving Methods in Automation Technology," NASA-TM-83216, May 1983.

13. James, J., "A Survey of Knowledge-Based Systems for Computer-Aided Control System Design," Proc. 1987 American Control Conference, Minneapolis, June 1987.

14. Kandel, E., and Schwartz, J., Principles of Neural Science, Elsevier Science Publishing Co., Inc., New York, 1985.

15. Hutchison, W., and Stephens, K., "Integration of Distributed and Symbolic Knowledge Representations," Proc. First Internat. Conf. on Neural Networks, San Diego, June 1987.

16. Buchanan, B., and Shortliffe, E., Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project, Addison-Wesley Publishing Company, Reading, Mass., 1984.

17. Birdwell, J., Cockett, J., and Gabriel, J., "Domains of Artificial Intelligence Relevant to Systems," Proc. 1986 American Control Conference, Seattle, June 1986.

18. Heudin, J., Burg, B., and Zavidovique, B., "A Flexible Operating System for Real-Time Process Control," Applications of Artificial Intelligence IV, Proc. SPIE, Vol. 657, April 1986.

19. Handelman, D., and Stengel, R., "An Architecture for Real-Time Rule-Based Control," Proc. 1987 American Control Conference, Minneapolis, June 1987.

20. Handelman, D., and Stengel, R., "Combining Expert System and Analytical Redundancy Concepts for Fault-Tolerant Flight Control," AIAA J. Guidance, Control, and Dynamics, Vol. 11 (to appear), 1988.

21. Gilmore, J., and Gingher, K., "A Survey of Diagnostic Expert Systems," Applications of Artificial Intelligence V, Proc. SPIE, Vol. 786, May 1987.

22. Tong, R., "A Control Engineering Review of Fuzzy Systems," Automatica, Vol. 13, No. 6, Nov. 1977.

PERSPECTIVES ON THE USE OF RULE-BASED CONTROL

David A. Handelman and Robert F. Stengel
Princeton University
Department of Mechanical & Aerospace Engineering
Princeton, New Jersey

Abstract. This paper addresses issues regarding the application of artificial intelligence techniques to real-time control. Advantages associated with knowledge-based programming are discussed. A proposed rule-based control technique is summarized and applied to the problem of automated aircraft emergency procedure execution. Although emergency procedures are by definition predominately procedural, their numerous evaluation and decision points make a declarative representation of the knowledge they encode highly attractive, resulting in an organized and easily maintained software hierarchy. Simulation results demonstrate that real-time performance can be obtained using a microprocessor-based controller. It is concluded that a rule-based control system design approach may prove more useful than conventional methods under certain circumstances, and that declarative rules with embedded procedural code provide a sound basis for the construction of complex, yet economical, control systems.

Keywords. Aerospace control; artificial intelligence; real-time operation; programming languages; hierarchical systems; man-machine systems; system failure and recovery.

INTRODUCTION

The apparent success of knowledge-based systems, such as expert systems, to provide a limited human-like decision-making capability within a well-defined problem domain gives strong support to their use in the design and implementation of complex control systems [1-4]. Before knowledge-based control techniques are widely accepted, however, many questions regarding their utility must be addressed. Why should a control system designer consider using knowledge-based programming techniques? What distinguishes knowledge-based techniques from conventional ones, how should various knowledge-based control ideas be evaluated and compared, which control problems call for a knowledge-based solution, and what specific benefits result from their use?

This paper attempts to answer some of these questions. The first section provides a comparison between conventional and knowledge-based programming, outlines some advantages associated with knowledge-based systems, and identifies some of the difficulties involved with applying these techniques to time-critical control. The second section summarizes an on-going research effort in real-time knowledge-based control designed to overcome these difficulties. The third section illustrates the utility of the proposed control technique by applying it to the problem of automated aircraft emergency procedure execution, and the final section summarizes benefits associated with its use.

THE PROMISE OF KNOWLEDGE-BASED PROBLEM SOLVING

One way knowledge-based systems can be distinguished from conventional software is by the manner in which data and the routines used to manipulate data remain separated within the program. As opposed to being written in a procedural manner whereby syntactic restrictions dictate an intermixing of code and data, knowledge-based systems use declarative statements, often in the form of rules, to declare and associate pieces of data. As the system runs, modifications and additions to the data are obtained through the use of an inference engine. In this case, the concept of data is generalized to mean knowledge, and the inference engine acts on the knowledge base (previously recognized as the data base) in hopes of inferring additional knowledge from that which already exists. The solution to a problem addressed by a knowledge-based system can come in the form of the additional knowledge (or information) inferred as a result of the system's execution, as well as in the form of sequenced actions performed as side effects of its execution.

The fundamental separation of a knowledge-based system into inference engine and knowledge base results in an enhanced capability for decision making and subsequent problem solving. The value of this enhancement, however, can mean different things to different people. While the user of a knowledge-based system may be impressed by the performance of the program itself, the designer of such a program also will appreciate the convenient environment provided by the adoption of knowledge-based system techniques. For example, although creation of an effective and consistent knowledge base is the toughest part of system construction, the mechanical ease with which it can be prototyped, tested, and modified (given the proper software and hardware tools) reflects a significant increase in programmer productivity in comparison to similar programming efforts based on conventional methods [5,6]. The benefits of using knowledge-based techniques, therefore, include not only what the resulting program can do, but also the efficient manner in which such a program may be created. It is within this context -- increased programmer productivity as well as program performance -- that the utility of knowledge-based techniques should be evaluated.

Many factors complicate the use of knowledge-based systems technology in control. Time-critical and numeric in nature, conventional control algorithms exhibit computational characteristics radically different from those exhibited by knowledge-based systems. The strength of knowledge-based systems comes from a symbolic processing capability, i.e., the ability to reason with non-numeric data. Unfortunately, symbolic computation facilities, in addition to being monetarily expensive, traditionally

Presented at the IFAC Workshop on
Artificial Intelligence in Real-Time Control,
Swansea, UK, September 1988.

result in slow execution speeds (in comparison to numeric facilities) and weak support for the representation and manipulation of floating-point numeric data types. Under the assumption that knowledge-based control efforts should build upon, and not attempt to replace, existing effective numerical control algorithms, a major challenge to control system designers is to integrate efficiently symbolic and numeric computation in a real-time environment.

Most control-oriented real-time knowledge-based systems developed to date can be characterized by (1) a separation within the control system of the symbolic and numeric processing environments (software and/or hardware), and (2) a supervisory role for the knowledge-based system, usually involving monitoring, diagnosis, and planning. Separation of the symbolic and numeric processing environments is justified by the fact that specialized software and hardware exist for this purpose. Moreover, as indicated in Fig. 1, sensing and control functions, conventionally based on numerical algorithms, usually are considered numeric processing tasks, whereas more general control system information processing usually is considered symbolic. Numeric processing historically has demonstrated the capability for high-bandwidth operation, whereas symbolic processing typically has been associated with low-bandwidth operation. Because sensing and control functions impose strict time constraints on control system design, separation of these high-bandwidth operations from the conventionally lower-bandwidth information processing operations seems essential.

Although justified, the separation of symbolic and numeric processing within a control system can limit severely the interaction between processing environments and the throughput of the system as a whole [7,8]. In time-critical applications such as fault-tolerant flight control, the throughput of a knowledge-based control system must be high. The present research goal is to integrate the desirable attributes of procedural and declarative techniques for efficient and effective control system programming, combining convenience in design with speed, economy, and high symbolic/numeric integration in implementation.

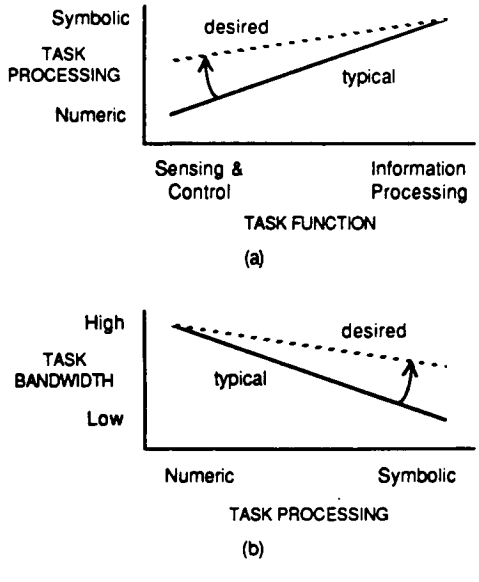


Fig. 1. Characteristic Relationship between Task Function, Processing, and Bandwidth

RULE-BASED CONTROL

Princeton Rule-Based Controller (PRBC) [9-11] embodies a control technique whereby control actions occur as a consequence of search through a knowledge base of parameters, rules, and procedures. Parameters, each with an assigned value, collectively represent a partial description of the "state of the world" pertinent to a given control objective. Search implies the attempt to increase the system's knowledge of the world by inferring additional parameter values, most of which are assumed unknown when the search begins. Information expressing relationships and dependencies between parameters is contained in rules, each of which contains a premise and an action. If a rule premise is true when tested, its action is executed, possibly causing the inference of additional parameter values. Rule testing is guided by an inference engine.

Control actions occur as side effects of search. Buried within the premises and actions of rules are procedures (or calls to procedures) that invoke time-critical control tasks that are best achieved using proven analytical techniques. These procedures are treated as building blocks upon which higher-level control actions are built using rules. Thus, in its simplest form, rule-based search conveniently implements deeply nested IF-THEN-ELSE clauses. The resultant ability to perform complex conditional branching in an organized manner provides a convenient mechanism with which to manipulate analytically derived numerical procedures. In this scenario, it is not the end result of a search that is important to control system operation but the side effects of its execution.

Such use of rules for control results in a tight functional integration between symbolic rules and numerical procedures. In this sense, the technique addresses the desired movement of the curve represented in Fig. 1a by admitting symbolic processing at the sensing and control level. Figure 1b, however, implies a concurrent desire to increase the bandwidth of symbolic processing, thereby addressing the issue of system throughput. The approach adopted here involves two distinct design phases. Phase I involves the development of rules and procedures in separate environments. LISP is used to create and test rules, whereas Pascal is used to derive procedures. Phase II involves final rule and procedure debugging in an integrated Pascal implementation environment. Phase I utilizes a LISP-based inference engine, whereas Phase II uses one based in Pascal. The transition from Phase I to Phase II involves the automatic translation of the LISP-based parameters, rules, and parameter-rule association information into a form acceptable to the Pascal-based inference engine. This knowledge base translation represents a form of automatic code optimization, and results in a compile-ready program implementing a search environment functionally similar to that based in LISP, but exhibiting the proven real-time control system performance characteristics of Pascal [12].

The advantages associated with an integrated Pascal implementation environment are many. In addition to increasing search speed dramatically, knowledge base translation allows rules and procedures to communicate through common data structures. Because parameters are implemented as Pascal variables, their values can be inspected and modified easily by procedures. Similarly, rules are free to access variables other than knowledge base parameters, such as the elements of a matrix routinely used by a numerical procedure. Finally, additional integration results from the ability to place arbitrary Pascal statements within the premises and actions of rules. The search process can thereby invoke timely pieces of Pascal code (calling a procedure, for example) as easily as Pascal code can invoke the process of search.

Although the search technique utilized here for control was inspired by that used in rule-based expert systems, it differs significantly in both implementation and intent. The use of rules in expert systems often is influenced by the characteristics of pure production systems. In production systems, each rule is considered a distinct and modular piece of knowledge, with explicit dependence of one rule on another being discouraged. Rules are meant to communicate with other rules through the indirect and limited link provided by the parameters of the knowledge base, but the intentional calling of one rule by another using parameter values as messages is looked upon with disfavor. As Davis and King [2] point out, "It is the premeditated nature of such message passing (typically in an attempt to 'produce a system with a specified behavior') that is the primary violation of the 'spirit' of production system methodology."

Given that the goal of a control system is, in fact, to produce a system with a specified behavior, it is not surprising that some control system designers view the application of pure production system techniques to control to be misguided. It is the intent of the rule-based control concept discussed here, however, not to take a pure production system approach, but to utilize the declarative power of the production system formalism for its beneficial effects. This effort may be summarized, therefore, as the attainment of procedural activity through the manipulation of declarative expressions, which is not an entirely new idea [13]. The uniqueness in this effort comes from its application to the design and implementation of time-critical control systems, particularly flight control systems. Although violating the spirit of pure production systems in order to provide the specified behavior demanded for control, the proposed control technique retains many advantages related to production systems. Davis and King [2] provide an interesting perspective related to program performance and programmer productivity as mentioned above.

> Since it is possible to imagine coding any given Turing machine in either procedural or production system terms, in the formal sense their computational power is equivalent. This suggests that, given sufficient effort, they are ultimately capable of solving the same problems. The issues we wish to examine are not, however, questions of absolute computational power but the impact of a particular methodology on program structure, as well as of the relative ease or difficulty with which certain capabilities can be achieved.

The next section illustrates the straightforward manner in which significant control system capabilities may be obtained using rules.

DECLARATIVE EXECUTION OF AIRCRAFT EMERGENCY PROCEDURES

The rule-based control technique described above was developed as part of a research program in applications of artificial intelligence theory to fault-tolerant flight control [14]. Presently, the technique is being applied to the automation of aircraft emergency procedures. Aircraft emergencies require a quick response and relatively complex decision making by a flight crew. Prescribed emergency procedures contained in the Pilot's Operating Handbook are designed to guide the flight crew through the decision-making process. Although these emergency procedures are by definition predominately procedural, their numerous evaluation and decision points make a declarative representation of the knowledge they encode highly attractive.

Consider as a simple example the following excerpt of an emergency procedure associated with the electrical system of the CH-47C tandem-rotor helicopter [15]:

> 4-96. FAILURE OF ONE AC GENERATOR.
>
> 4-97. Should one ac generator fail, the remaining ac generator will assume the entire load under normal conditions. This condition will be noticed by the illumination of a generator caution light and by a zero indication on the ac loadmeter for that generator. Attempt to place the inoperative ac generator into operation by performing the following:
>
> a. Master caution lights - PUSH TO RESET.
>
> b. All circuit breakers - CHECK.
>
> c. Generator switch - TEST, then RESET, then ON. If the generator is inoperative, move the generator switch to OFF. (Move the generator switch to TEST and observe the generator caution light. If the caution light goes out, the generator is delivering proper voltage and frequency and a short circuit on a bus is indicated. If the caution light remains on, the generator is inoperative.)

The following discussion highlights how this emergency procedure (with reference to Generator No.1 only) can be made to execute automatically as a side effect of search through a knowledge base of parameters and rules. The inference engine executes on the highest level a repetitive cycle involving knowledge base initialization and goal-directed search on a parameter named PROC_SEARCH_COMPLETED. This is represented in Pascal as

```
repeat
   initialize_knowledge_base;
   determine_value_of(PROC_SEARCH_COMPLETED)
until false;
```

Knowledge base initialization assigns the value of UNKNOWN to all parameters without a stored "initial value". Goal-directed (backward-chaining) search on parameter PROC_SEARCH_COMPLETED results in the inference engine testing rules capable of supplying a value for this parameter in their action, such as the following.

```
[RULE_PROC_1
  [PREMISE
    '($OR ($EQ PROC_REQUIRED      'FALSE)
          ($EQ PROC_STEP_CHECKED 'TRUE]
  [ACTION
    '($SETQ PROC_SEARCH_COMPLETED 'TRUE]]
```

Note that the premise of this rule depends on other parameters whose values also are as yet unknown (parameters receiving a value of UNKNOWN during knowledge base initialization are written entirely in capital letters). When tested, the premise first searches for a value for parameter PROC_REQUIRED. A value of FALSE implies that the emergency procedure is not required (the conditions under which it is to be executed do not exist). The following rules determine whether or not procedure requirements are met.

```
[RULE_PROC_REQ_1
  [PREMISE
   '($AND ($EQ Proc_Step 0)
          ($EQ Gen_1_Control_Switch_Status 'ON)
          ($EQ Gen_1_Loadmeter_Status     0)
          ($EQ Gen_1_Caution_Light_Status 'ON)
          ($EQ Gen_2_Caution_Light_Status 'OFF]
  [ACTION
    '(($SETQ Proc_Step 1)
      ($SETQ PROC_REQUIRED 'TRUE]]
[RULE_PROC_REQ_2
  [PREMISE
    '($GT Proc_Step 0)
  [ACTION
    '($SETQ PROC_REQUIRED 'TRUE]]

[RULE_PROC_REQ_3
  [PREMISE
    '($EQ Rule_Tested 'TRUE]
  [ACTION
    '($SETQ PROC_REQUIRED 'FALSE]]
```

Rules are tested in order of appearance in the knowledge base until a value for the specified parameter is obtained. These three rules are tested in order when a value for PROC_REQUIRED is needed within the premise of RULE_PROC_1. The premise of the first rule holds if the procedure is not already executing (the active procedure step number, represented by parameter Proc_Step, is 0) and the conditions specified by the remainder of the $AND clause prevail. The second rule holds if the procedure is already executing (Proc_Step is greater than 0), and the third rule always holds whenever tested. Thus, if during search only the third rule holds, then PROC_REQUIRED obtains a value of FALSE, and PROC_SEARCH_COMPLETED is assigned a value of TRUE within the action of RULE_PROC_1 quickly ending the search cycle. However, if RULE_PROC_REQ_1 holds, then the step number is set to 1, PROC_REQUIRED obtains a value of TRUE, the first subclause of RULE_PROC_1 fails, and a goal-directed search for PROC_STEP_CHECKED begins. A similar search will prevail if RULE_PROC_REQ_1 fails, but RULE_PROC_REQ_2 holds.

Emergency procedures are executed here as a series of steps. Steps are executed one at a time, in order. Each step in general involves an initial action (invoked for Step 1 by parameter PROC_STEP_1_STARTED), a time delay (reflected by parameter Proc_Step_Delay) over which the initial action is allowed to take effect, and a final action (invoked for Step 1 by parameter PROC_STEP_1_FINISHED). For example, an emergency procedure may require the flipping of a switch and the monitoring of its effect. The time delay between the initial and final step actions gives the physical system being probed a chance to react. When a final step action is taken, Proc_Step is incremented, allowing the next emergency procedure step to be performed.

Rules invoked by cyclic goal-directed search effectively encode the logic required to execute these emergency procedure steps with the appropriate time delays. As shown by the rules above, cyclic search on parameter PROC_SEARCH_COMPLETED results in cyclic search on parameter PROC_STEP_CHECKED (assuming that PROC_REQUIRED is TRUE). In effect, the control system continually asks the question, "Has the appropriate emergency procedure step been checked?" Rules capable of answering this question for Step 1 are shown below.

```
[RULE_PROC_STEP_1A
  [PREMISE
    '($AND ($EQ Proc_Step 1)
           ($EQ Proc_Step_Timer     'STOPPED)
           ($EQ PROC_STEP_TIMER_SET 'TRUE)
           ($EQ PROC_STEP_1_STARTED 'TRUE]]
  [ACTION
    '($SETQ PROC_STEP_CHECKED 'TRUE]]

[RULE_PROC_STEP_1B
  [PREMISE
    '($AND
      ($EQ Proc_Step 1)
      ($EQ Proc_Step_Timer 'RUNNING)
      ($OR [$AND
             ($EQ PROC_STEP_1_BYPASSED 'TRUE)
             ($EQ PROC_STEP_TIMER_SET  'TRUE]
           ($EQ PROC_STEP_TIMEOUT 'FALSE]]
  [ACTION
    '($SETQ PROC_STEP_CHECKED 'TRUE]]

[RULE_PROC_STEP_1C
  [PREMISE
    '($AND ($EQ Proc_Step 1)
           ($EQ Proc_Step_Timer      'RUNNING)
           ($EQ PROC_STEP_TIMEOUT    'TRUE)
           ($EQ PROC_STEP_1_FINISHED 'TRUE)
           ($EQ PROC_STEP_TIMER_SET  'TRUE]]
  [ACTION
    '($SETQ PROC_STEP_CHECKED 'TRUE]]
```

When a value for PROC_STEP_CHECKED is required by RULE_PROC_1, these rules are tested. Each rule performs a distinct intra-step function, monitoring the "state" of Procedure Step 1 and quickly performing a piece of the step if required. At least one rule always holds. Within the first rule, the parameter Proc_Step_Timer is checked for a value of STOPPED. This indicates that the step has not begun, in which case the timer is started (via search for parameter PROC_STEP_TIMER_SET using rules not shown) and the initial step action is taken (via search for parameter PROC_STEP_1_STARTED). If the first rule fails (Proc_Step_Timer is RUNNING), a search for parameter PROC_STEP_1_BYPASSED within the second rule checks whether or not the final step action already has been performed by the flight crew. If it has, PROC_STEP_1_BYPASSED will return a value of TRUE and a search on PROC_STEP_TIMER_SET will stop the timer and increment Proc_Step. If PROC_STEP_1_BYPASSED is FALSE, the next premise subclause tests whether or not the required step time delay has been achieved. If PROC_STEP_TIMEOUT is FALSE (inferred with a rule not shown), the rule premise holds without performing any emergency procedure actions. If PROC_STEP_TIMEOUT is TRUE, the third rule performs the final step action (via search for parameter PROC_STEP_1_FINISHED) and stops the timer and increments Proc_Step (via search for parameter PROC_STEP_TIMER_SET).

Thus, each search for parameter PROC_STEP_CHECKED performs a single piece of a single procedure step. Execution of the entire emergency procedure requires many search cycles. The following three Step 1 rules encode the first part of the emergency procedure excerpt shown above.

```
[RULE_PROC_STEP_1_ST
  [PREMISE
    '($AND ($EQ Advisory_Mode 'ON)
           ($PASCAL "advise
             ('*** EMERGENCY PROCEDURE ***
               Failure of One AC Generator:
               a. Master caution light -
                  PUSH TO RESET');")
           ($SETQ Proc_Step_Delay 1.0]
  [ACTION
    '($SETQ PROC_STEP_1_STARTED 'TRUE]]

[RULE_PROC_STEP_1_BY
  [PREMISE
    '($EQ Master_Caution_Light_Status 'OFF]
  [ACTION
    '($SETQ PROC_STEP_1_BYPASSED 'TRUE]]

[RULE_PROC_STEP_1_FIN
  [PREMISE
   '($AND
     ($EQ   Operational_Mode           'ACTIVE)
     ($SETQ Master_Caution_Light_Command 'PUSH]
  [ACTION
    '($SETQ PROC_STEP_1_FINISHED 'TRUE]]
```

The premise of the first rule contains three subclauses. The first subclause holds if parameter Advisory_Mode has value ON. If so, the next two subclauses are evaluated. The $PASCAL subclause contains embedded Pascal code that calls a procedure which sends an advisory message to the cockpit for evaluation by the flight crew. The $SETQ subclause assigns to parameter Proc_Step_Delay a value of 1. Both of these subclause operators, as members of a Boolean premise, always return TRUE. Similarly, the premise of the third rule generates a master caution light command within the $SETQ subclause if parameter Operational_Mode has value ACTIVE. Thus, for Procedure Step 1, if Advisory_Mode is ON and Operational_Mode is ACTIVE, the control system will reset the master caution light if after 1 sec this task has not been performed by the flight crew. In general, the parameter Advisory_Mode can have values ON and OFF, and parameter Operational_Mode can have values ACTIVE and STANDBY. If Advisory_Mode is OFF, no recommendations are sent to the flight crew, and advisory delays are eliminated. If Operational_Mode is STANDBY, actions may be recommended but not executed by the flight control system.

If any of these three rules fails, an additional rule supplying an appropriate "default" value for the required parameter is tested, such as the following.

```
[RULE_PROC_STEP_1_BY_0
  [PREMISE
    '($EQ Rule_Tested 'TRUE]
  [ACTION
    '($SETQ PROC_STEP_1_BYPASSED 'FALSE]]
```

The remaining parts of the emergency procedure excerpt given above are implemented in a similar fashion. Consider the next three rules.

```
[RULE_PROC_STEP_2_ST
  [PREMISE
    '($AND
       ($EQ Advisory_Mode 'ON)
       ($PASCAL "advise
         ('b. All circuit breakers - CHECK');")
       ($SETQ Proc_Step_Delay 2.0]
  [ACTION
    '($SETQ PROC_STEP_2_STARTED 'TRUE]]

[RULE_PROC_STEP_2_BY
  [PREMISE
    '($EQ BREAKERS_CHECKED_BY_FLIGHTCREW 'TRUE]
  [ACTION
    '($SETQ PROC_STEP_2_BYPASSED 'TRUE]]

[RULE_PROC_STEP_2_FIN
  [PREMISE
    '($AND ($EQ OPERATIONAL_MODE      'ACTIVE)
           ($EQ BREAKERS_CHECKED_BY_FCS 'TRUE]
  [ACTION
    '($SETQ PROC_STEP_2_FINISHED 'TRUE]]
```

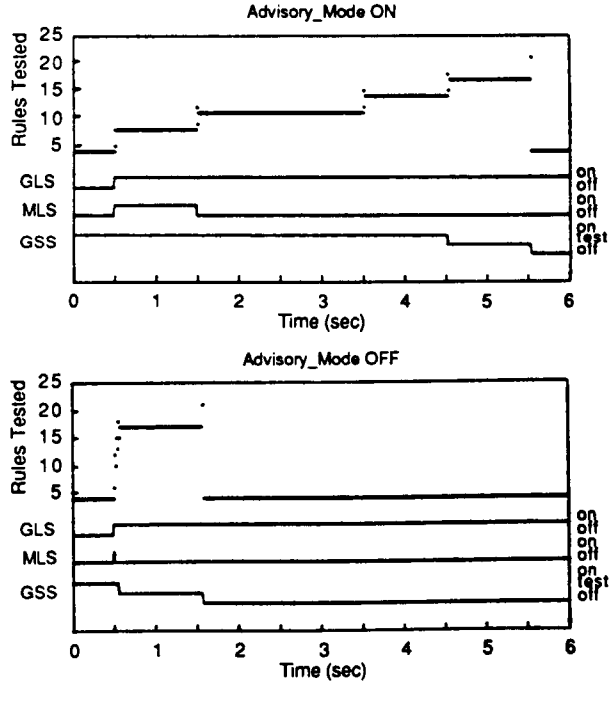
The first rule provides an advisory with a 2 sec grace period. Within this time period, a goal-directed search on parameter BREAKERS_CHECKED_BY_FLIGHTCREW performed within the second rule invokes additional rules that check if all circuit breakers have been checked by the flight crew. The third rule invokes other rules performing this task automatically if and when necessary.

The last part of the emergency procedure is implemented with two steps. Rules of Procedure Step 3 verify movement of the generator switch to the TEST position. If Advisory_Mode is OFF, the control system performs the action immediately. If Advisory_Mode is ON, a maximum advisory delay of 1 sec is tolerated. Finally, rules of Procedure Step 4 wait one more second (regardless of Advisory_Mode) before moving the generator switch to ON or OFF, depending on the status of the generator caution light.

The result of using a declarative rule-based representation for emergency procedure execution is an organized and easily maintained software hierarchy. Additionally, by using cyclic search with "time-sliced" procedure steps, the action of the control system can be made to emulate a multi-tasking operating system. This capability is demonstrated below.

Figure 2 shows a time history of the amount of "effort" expended by a single-processor rule-based controller executing the emergency procedure described above. The data was obtained with the Princeton Rule-Based Controller Development System [9], employing a commercially-available single-board computer outfitted with an 8-MHz 80286 processor and an 8-MHz 80287 math coprocessor. The controller knowledge base contained 34 parameters and 34 rules. Off-line LISP-to-Pascal knowledge base translation required 8.6 min on a personal computer functionally similar to the PRBC processor described above. Pascal representation of the knowledge base required 16 KBytes of Random-Access Memory (some as code, some as data), and the inference engine required 13 KBytes of code.

During simulation runs, certain parameters were given a fixed initial value: Gen_1_Loadmeter_Status was 0, Gen_2_Caution_Light_Status was OFF, BREAKERS_CHECKED_BY_FLIGHTCREW was FALSE, BREAKERS_CHECKED_BY_FCS was TRUE. Emergency procedure execution was triggered by changing the value of Gen_1_Caution_Light_Status from OFF to ON. Each data point in the top plots of Fig. 2 corresponds to the number of rules tested by the inference engine during a goal-directed search cycle. Adjacent data points are separated by the amount of time required to complete a search cycle. Figure 2 shows that although the value of the parameter Advisory_Mode has a large effect on emergency procedure step timing (as intended), the inference engine consistently executes search cycles at a very high rate.



GLS = Gen_1_Caution_Light_Status
MLS = Master_Caution_Light_Status
GSS = Gen_1_Control_Switch_Status

Fig. 2. Execution of Single Emergency Procedure

Figure 3 shows a time history of rule testings performed by the same rule-based controller concurrently executing five copies of the same emergency procedure. Although all procedures trigger off the same generator caution light, each manipulates its own master caution light and generator switch, and employs a unique set of step delays. Figure 3 shows that this version of the rule-based controller, with a knowledge base of 154 parameters and 170 rules, still provides real-time multi-tasking performance with a single economical processor.

CONCLUSIONS

The main conclusion to be drawn is that a rule-based control design approach may prove more useful than conventional methods under certain circumstances, especially when complex decision making is required. The proposed rule-based control technique provides basic programming constructs required in real-time applications such as flight control. Capabilities including event scheduling, selection, and synchronization, as well as data passing and sharing, are implemented in an extremely flexible and modular representation. Consequently, declarative rules with embedded procedural code provide a sound basis for the construction of complex, yet economical, control systems.

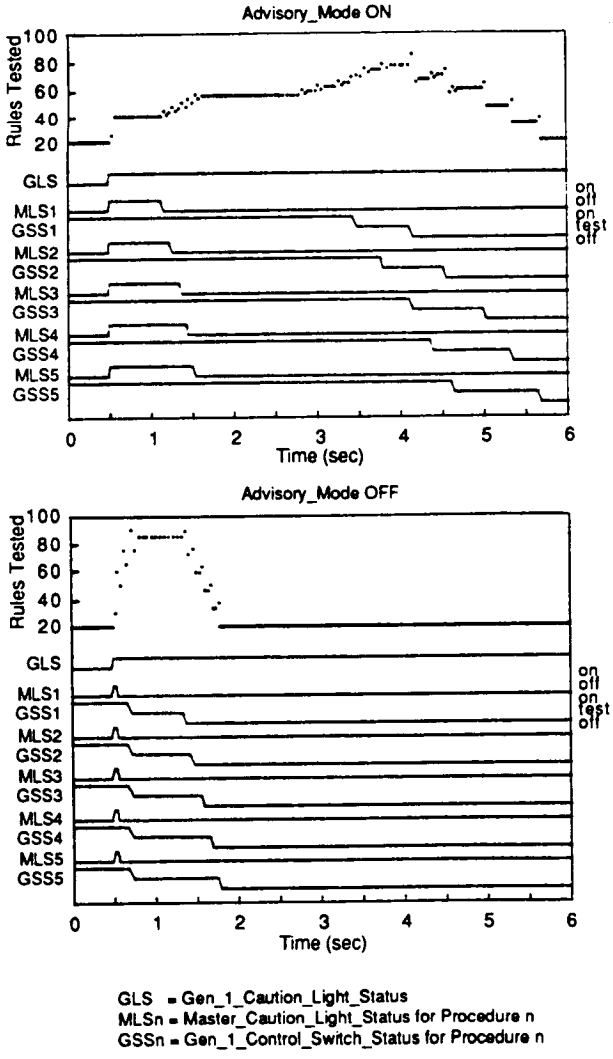


Fig. 3. Execution of Multiple Concurrent Emergency Procedures

ACKNOWLEDGMENT

This project was sponsored by the U. S. Army Research Office under Contract No. DAAG29-84-K-0048.

REFERENCES

1. Barr, A., Cohen, P., and Feigenbaum, E., The Handbook of Artificial Intelligence, William Kaufmann, Los Altos, California, 1982.

2. Davis, R. and King, J. J. in Buchanan, B., and Shortliffe, E., Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project, Addison-Wesley Publishing Company, Reading, Mass., 1984.

3. Hankins, W., Pennington, J., and Barker, L., "Decision-Making and Problem-Solving Methods in Automation Technology," NASA-TM-83216, May 1983.

4. James, J., "A Survey of Knowledge-Based Systems for Computer-Aided Control System Design," Proc. 1987 American Control Conference, Minneapolis, June 1987.

5. Ricks, W., and Abbott, K., "Traditional Versus Rule-Based Programming Techniques: Application to the Control of Optional Flight Information," Applications of Artificial Intelligence V, Proc. SPIE, Vol. 786, May 1987.

6. Hueschen, R., and McManus, J., "Application of AI Methods to Aircraft Guidance and Control," Proc. 1988 American Control Conference, Atlanta, June 1988.

7. O'Reilly, C., and Cromarty, A., "'Fast' is not 'Real-Time': Designing Effective Real-Time AI Systems," Applications of Artificial Intelligence II, Proc. SPIE 548, 1985.

8. Birdwell, J., Cockett, J., and Gabriel, J., "Domains of Artificial Intelligence Relevant to Systems," Proc. 1986 American Control Conference, Seattle, June 1986.

9. Handelman, D., and Stengel, R., "An Architecture for Real-Time Rule-Based Control," Proc. 1987 American Control Conference, Minneapolis, June 1987.

10. Handelman, D., and Stengel, R., "Combining Expert System and Analytical Redundancy Concepts for Fault-Tolerant Flight Control," AIAA J. Guidance, Control, and Dynamics, Vol. 11, Jan-Feb, 1989.

11. Handelman, D., and Stengel, R., "Rule-Based Mechanisms of Learning for Intelligent Adaptive Flight Control," Proc. 1988 American Control Conference, Atlanta, June 1988.

12. Westermeier, T., and Hansen, H., "The Use of High Order Languages in Microprocessor-Based Systems," Proc. 1984 American Control Conference, San Diego, June 1984.

13. Munakata, T., "Procedurally Oriented Programming Techniques in Prolog," IEEE Expert, Vol. 1, No. 2, Summer 1986.

14. Stengel, R., "Artificial Intelligence Theory and Reconfigurable Control Systems," Department of Mechanical and Aerospace Engineering Report 1664, Princeton University, Princeton, New Jersey, June 1984.

15. "Operator's Manual: Army Model CH-47B and CH-47C Helicopters," Army Technical Manual No. 55-1520-227-10, Aug. 1970.

STOCHASTIC ROBUSTNESS OF LINEAR CONTROL SYSTEMS

Robert F. Stengel and Laura E. Ryan
Department of Mechanical and Aerospace Engineering
Princeton University

ABSTRACT

A simple numerical procedure for estimating the *stochastic robustness* of a linear, time-invariant system is described. Monte Carlo evaluation of the system's eigenvalues allows the *probability of instability* and the related *stochastic root locus* to be estimated. This definition of robustness is an alternative to existing deterministic definitions that address both structured and unstructured parameter variations directly. This analysis approach treats not only Gaussian parameter uncertainties but non-Gaussian cases, including uncertain-but-bounded variations. Trivial extensions of the procedure admit alternate discriminants to be considered. Thus, the probabilities that stipulated degrees of instability will be exceeded or that closed-loop roots will leave desirable regions also can be estimated. Results are particularly amenable to graphical presentation.

INTRODUCTION

Control system robustness is defined as the ability to maintain satisfactory stability or performance characteristics in the presence of all conceivable system parameter variations. While assured robustness may be viewed as an alternative to gain adaptation or scheduling to accommodate known parameter variations, more often it is seen as protection against uncertainties in plant specification. Consequently, a statistical description of control system robustness is consistent with what may be known about the structure and parameters of the plant's dynamic model.

Guaranteeing robustness has long been a design objective of control system analysis, although in most instances, insensitivity to parameter variations has been treated as a deterministic problem (see Ref. 1 for a comprehensive presentation of both classical and modern robust control). Stability (gain and phase) margins are useful concepts for designing robust single-input/single-output systems, addressing disturbance rejection and other performance goals in the process, and they are amenable to the manual graphical procedures that preceded the widespread use of computers. With the help of these computers, singular-value analysis has extended the frequency-domain approach to multi-input/multi-output systems (e.g., [2,3]); however, guaranteed-stability-bound estimates often are unduly conservative, and the relationship to parameter variations in the physical system is weak. Structured-singular-value analysis [4] reduces this conservatism somewhat, and alternate treatments of structured parameter variations have been proposed (e.g., [5-7]), though these approaches remain deterministic. Elements of stochastic stability [8] have application to robustness but have yet to be presented in that context.

Presented at the 1989 Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, MD.

The notion of *probability of instability*, which is central to the analysis of stochastic robustness, was introduced in Ref. 9, with application to the robustness of the Space Shuttle's flight control system, and it is further described in Ref. 10. This method determines the *stochastic robustness* of a linear, time-invariant system by the probability distributions of closed-loop eigenvalues, given the statistics of the variable parameters in the plant's dynamic model. The probability that all of these eigenvalues lie in the open left-half s plane is the scalar measure of robustness.

With the advent of fast graphics workstations and supercomputers, the stochastic robustness of a system is easily computed by Monte Carlo simulation, and results can be displayed pictorially, providing insight into otherwise hidden robustness properties of the system. The method is computationally simple, requiring only matrix manipulation and eigenvalue computation, and it is inherently non-conservative, given a large enough sample space. Furthermore, the analysis of stochastic robustness is a logical adjunct to parameter-space control design methods [11-14]. Details of the approach and examples are given in the sequel.

PROBABILITY OF INSTABILITY

Consider a linear, time-invariant (LTI) system of the form,

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{p})\mathbf{x}(t) + \mathbf{G}(\mathbf{p})\mathbf{u}(t) \qquad (1$$

$$\mathbf{y}(t) = \mathbf{H}(\mathbf{p})\mathbf{x}(t) \qquad (2$$

where $\mathbf{x}(t)$, $\mathbf{u}(t)$, $\mathbf{y}(t)$, and $\mathbf{p}$ are state, control, output, and parameter vectors of dimension n, m, q, and r, respectively, accompanied by conformable dynamic, control, and output matrices $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$, which may be arbitrary functions of $\mathbf{p}$. The plant is subject to LTI control,

$$\mathbf{u}(t) = \mathbf{u}_C(t) - \mathbf{C}\mathbf{H}(\mathbf{p})\mathbf{x}(t) \qquad (3$$

$\mathbf{u}_C(t)$ is a command input vector, and, for simplicity, the (m x n) control gain matrix $\mathbf{C}$ is assumed to be known without error. The n eigenvalues, $\lambda_i = \sigma_i + j\omega_i$, $i = 1$ *to* n, of the matrix $[\mathbf{F}(\mathbf{p}) - \mathbf{G}(\mathbf{p})\mathbf{C}\mathbf{H}(\mathbf{p})]$ determine closed-loop stability and can be determined as the roots of the determinant equation,

$$|s\mathbf{I}_n - [\mathbf{F}(\mathbf{p}) - \mathbf{G}(\mathbf{p})\mathbf{C}\mathbf{H}(\mathbf{p})]| = 0 \qquad (4$$

where s is a complex operator and $\mathbf{I}_n$ is the (n x n) identity matrix. System stability requires that no eigenvalues have positive real parts. The relationship between parameters and eigenvalues is complicated.

Even if **F**, **G**, and **H** are linear functions of **p**, the associations between **p** and the λ_i are nonlinear, and there is the further possibility of products of parameters in the feedback term. Consequently, while small-parameter-variation sensitivities of the eigenvalues can be estimated by linear methods [10], the hopes for generally applicable analytic expressions are slim.

Putting aside the mathematical intricacies, note that the probability of stability plus the probability of instability is one:

$$\Pr(\text{stability}) + \Pr(\text{instability}) = 1 \qquad (5$$

Stochastic robustness is achieved when the probability of stability (instability) is large (small). Since stability requires all the roots to be in the open-left-half s plane, while instability results from even a single right-half s plane root, we may write

$$\Pr(\text{instability}) \triangleq \mathbb{P} = 1 - \int_{-\infty}^{0} \text{pr}(\sigma)d\sigma \qquad (6$$

where σ is an n-vector of the real parts of the system's eigenvalues, pr(σ) is the joint probability density function of σ, and the integral that defines the probability of stability is evaluated over the space of individual components of σ.

Estimating the probability of stability of a closed-loop system from repeated eigenvalue calculation is a straightforward task. Denoting the probability density function of **p** as pr(**p**), eq. 4 is evaluated J times with each element of $\mathbf{p}_j$, j = 1 to J, specified by a random-number generator whose individual outputs are shaped by pr(**p**). This *Monte Carlo evaluation* of the probability of stability becomes increasingly precise as J becomes large. Then,

$$\int_{-\infty}^{0} \text{pr}(\sigma)d\sigma = \lim_{J \to \infty} \frac{N(\sigma_{\max} \le 0)}{J} \qquad (7$$

N(.) is the number of cases for which all elements of σ are less than or equal to zero, that is, for which $\sigma_{\max} \le 0$, where $\sigma_{\max}$ is the maximum real eigenvalue component in σ. An important feature of this definition is that it does not depend on the eigenvalues and eigenvectors retaining fixed structures. As parameters change, complex roots may coalesce to become real roots (or the reverse), and modes may exchange relative frequencies. The only matter for concern is whether or not all real parts of the eigenvalues remain in the left-half s plane. The stable space of σ is a hypercube with one corner at the origin and all other corners at various infinite points.

There is, of course, no limitation on admissible specifications for the multivariate pr(**p**): it may be Gaussian or non-Gaussian, as appropriate. Rayleigh, correlated, and any other well-posed distributions are admissible, the principal challenge being to properly shape (and correlate) the outputs of the random-number generator. In practice, system parameter uncertainties are most likely to be bounded, as typical quality control procedures eliminate out-of-tolerance devices, and there are physical limitations on component size, weight, shape, etc. The rectangular (uniform) distribution is particularly interesting, as it readily models bounded uncertainty, and it is the default distribution of most algorithms for random-number generation. Given binary distributions for each parameter, in which the elements of **p** take maximum or minimum values with equal probability, the Monte Carlo evaluation reduces to 2^r deterministic evaluations, the result is exact, and the probability associated with each possible value of **p** is $1/2^r$. Similarly, the distribution for r parameters, each of which takes w values (i.e., for quantized rectangular distributions), can be obtained from w^r evaluations; the probability of acquiring each value of **p** (for equally probable parameter values) is $1/w^r$.

Histograms and cumulative distributions for varying degrees of stability are readily given by the Monte Carlo estimate of $\int_{-\infty}^{\Sigma} \text{pr}(\sigma)d\sigma$, where Σ represents a maximum real eigenvalue component, and $-\infty < \Sigma < \infty$. The histogram is a plot of $\frac{N[(\Sigma - \Delta) < \sigma_{\max} \le \Sigma]}{J}$ vs. Σ; Δ is an increment in Σ, N[.] is the number of cases whose maximum real eigenvalue components lie in the increment, and J is the total number of evaluations. The histogram estimates the *stability probability density function*, pr(Σ), which is obtained in the limit for a continuous distribution of Σ as $\Delta \to 0$ and $J \to \infty$. The *cumulative probability distribution of stability*, Pr(Σ), is similarly estimated and presented as $\frac{N(\sigma_{\max} \le \Sigma)}{J}$ vs. Σ, the exact distribution being achieved in the limit as $J \to \infty$. Consequently,

$$\mathbb{P} = 1 - \Pr(0). \qquad (8$$

The regions of varying stability degree are hypercubes in σ space, each with one corner at the n-vector $\Sigma = [\Sigma\ \Sigma\ \Sigma\ \ldots\ \Sigma]^T$ and all remaining corners at appropriate infinite locations.

When has stochastic robustness been achieved? The answer is problem-dependent. In some applications involving bounded parameters, it will be possible to choose **C** such that $\mathbb{P} = 0$, and that is a desirable goal; however, if admissible parameter variations are unbounded, if **C** is constrained, or if the rank of **CH** is less than n, the minimum $\mathbb{P}$ may be greater than zero. **C** then must be chosen to satisfy performance goals and one of two robustness criteria: minimum $\mathbb{P}$, or $\mathbb{P}$ small enough to meet a reliability specification (e.g., one chance of instability in some large number of realizations).

STOCHASTIC ROOT LOCUS

While it is not necessary to plot the eigenvalues (or roots) of eq. 4 to determine or portray stochastic

robustness, stochastic root loci provide insight regarding the effects of parameter uncertainty on system stability. Consider, for example, a classical second-order system whose roots are solutions to the equation

$$s^2 + 2\zeta\omega_n s + \omega_n^2 = 0 \qquad (9$$

Suppose that the damping ratio (ζ) and natural frequency (ω_n) are nominally 0.707 and 1, respectively, and that each may be a Gaussian-distributed random variable with standard deviation of 0.2. Allowing first ζ to vary, then ω_n, 100-sample scatter plots of the roots are obtained (Fig. 1). These root loci are immediately recognized as following the classical configurations of root locus construction [15], with the heaviest density of roots in the vicinities of the nominal values. The density of roots depicts the likelihood that eigenvalues vary from their nominal values if either damping ratio or natural frequency is uncertain. These stochastic root loci include branches on the real axis and in the right-half *s* plane for large enough variations of ζ and ω_n.

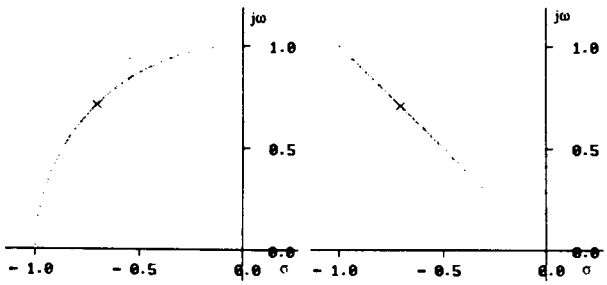


Figure 1. Stochastic root loci of a second-order system with Gaussian damping ratio or natural frequency. $\zeta_o = 0.707$, $\omega_{no} = 1$; 100 Monte Carlo evaluations.
a) Effect of ζ variation with 0.2 standard deviation.
b) Effect of ω_n variation with 0.2 standard deviation.

If both ζ and ω_n are uncertain and uncorrelated (i.e., $\mathbf{p} = [\zeta\ \omega_n]^T$), the scatter plots become "clouds" surrounding the nominal values; Fig. 2a is one representation of the resulting stochastic root locus based on the calculation of 4,000 samples. Further understanding can be gained by plotting the density of roots in a third dimension above the root locus plot. This is done in two steps. The first step is to divide the *s* plane into subspaces (or "bins"), as in Fig. 2b, and to count the number of roots in each bin as a sampled estimate of the *root density* ρ. The result is a multivariate histogram, with σ and ω serving as independent variables. Complex root bins are elemental areas, for which ρ_A is defined in units of roots/unit area. Real root bins are confined to the real axis; hence, ρ_L measures roots/unit length.

The second step is to portray the root density distribution. This can be done by brightening or darkening the bin outlines (Fig. 2b), graphing contours of equal root density on the two-dimensional plot, or by plotting an oblique view of the three-dimensional histogram or root density surface, as in Fig. 2c. The plotted surfaces would become smoother as the number of evaluations increased. Numerical smoothing also can be applied (judiciously) to account for sampling effects on the plotted surface. For this paper, the graphical presentations are relatively crude, but it is apparent that more sophisticated graphical processing, including the use of false color, hidden-line removal, surface generation, and shading can be applied to good effect. Root densities along the real axis present a special problem for 2-D presentation, in that their distributions are linear, not areal; oblique 3-D views provide a satisfactory alternative.

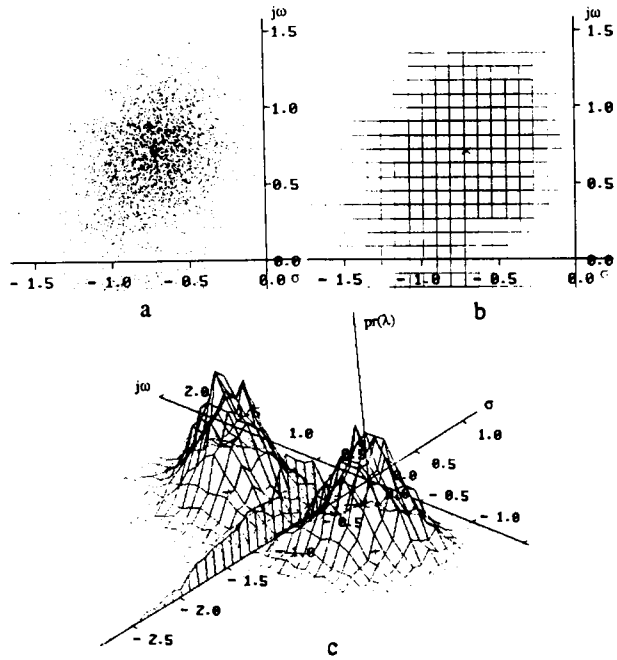


Figure 2. Stochastic root loci of a second-order system with Gaussian damping ratio and natural frequency. $\zeta_o = 0.707$, $\omega_{no} = 1$; 4,000 Monte Carlo evaluations.
a) Scatter plot.
b) 2-dimensional binned representation.
c) Oblique 3-dimensional representation.

There is an ostensible relationship between pr(Σ) and ρ; however, the relationship may be multivalued and ambiguous. When considering instability, distinction must be made between the number of *cases* with right-half-plane roots and the number of *roots* in the right-half plane. For example, a third-order system with a complex pair of roots and a real root can be unstable with 1, 2, or 3 roots in the right-half plane, yet *N* would be incremented by one in each case. A high-order system with real roots could be unstable with one or more roots in the same right-half-plane bin. Again, *N* would be incremented by one, although the bin's ρ depends upon the number of roots it contains.

A STOCHASTIC ROBUSTNESS EXAMPLE

Reference 16 provides a linear-quadratic-Gaussian (LQG) design problem with a closed-loop system that is nominally stable, but whose stability margins become vanishingly small as control and estimation gains become

large. That example is used here for a demonstration of stochastic stability robustness. An unstable second-order plant,

$$\begin{bmatrix}\dot{x}1\\ \dot{x}2\end{bmatrix}=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}x1\\ x2\end{bmatrix}+\begin{bmatrix}0\\ 1\end{bmatrix}u+\begin{bmatrix}1\\ 1\end{bmatrix}w \qquad (10$$

$$z=[1\ 0]\begin{bmatrix}x1\\ x2\end{bmatrix}+n \qquad (11,$$

is to be stabilized by an LQG regulator with controller cost function matrices,

$$\mathbf{Q}=Q\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},\quad \mathbf{R}=R=1 \qquad (12,13$$

and disturbance and measurement-error spectral density matrices

$$\mathbf{W}=W\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},\quad \mathbf{N}=N=1 \qquad (14,15$$

The corresponding LQG control and estimator gains, **C** and **K**, are [16]

$$\mathbf{C}=\left(2+\sqrt{4+Q}\right)[1\ 1]=[c\ c] \qquad (16$$

$$\mathbf{K}=\left(2+\sqrt{4+W}\right)[1\ 1]^T=[k\ k]^T \qquad (17$$

If the actual control effect matrix is $\mathbf{G}=[0\ \mu]^T$ rather than $[0\ 1]^T$, eq. 4 can be expressed for this problem (with the state consisting of the original state and its estimate, $\mathbf{x}^T=[x_1\ x_2\ \hat{x}_1\ \hat{x}_2]$), as,

$$\begin{vmatrix}(s-1) & -1 & 0 & 0\\ 0 & (s-1) & \mu c & \mu c\\ -k & 0 & (s-1+k) & -1\\ -k & 0 & (c+k) & (s-1+c)\end{vmatrix}=s^4+c_3s^3+c_2s^2+c_1s+c_0=0 \qquad (18$$

Using Routh's criterion, Doyle showed that μ remaining in $(a,b)=\{(1+1/ck),[1-(k+c-4)/2ck]\}$ is a necessary condition for stability to be retained [17].

Consider two cases with different LQG gains. In Case 1, c = k = 4 (the limiting case as Q and W approach zero), and in Case 2, c = k = 100. Because

$$c_1=k+c-4+2(\mu-1)ck,\quad c_0=1+(1-\mu)ck \qquad (19,20$$

the characteristic polynomial can be expressed as

$$s^4+c_3s^3+c_2s^2+(k+c-4-2ck)s+(1+ck)+\mu ck(2s-1)=0 \qquad (21$$

which is a root-locus problem with μck taken as the gain. The nominal roots are found with $\mu = 1$, and they are

Case 1 $\quad \lambda_{1-4}=-1,-1,-1,-1 \qquad (22$

Case 2 $\quad \lambda_{1-4}=-0.01,-0.01,-98,-98 \qquad (23$

Three features are immediately evident. The root locus gain is proportional to ck; hence, μ has a greater effect on the root locus in the latter case. There is a transmission zero at +1/2 that will draw one root into the right half plane. The excess of poles over zeros is three, indicating that additional instability must occur for large magnitudes of $(\mu - 1)$. There will be either one or two unstable roots among those going to infinity, depending on the sign of $(\mu - 1)$.

The stochastic root locus plots based on 3,500 Monte Carlo evaluations with $\mathbf{p}=\mu$ corroborate these predictions (Fig. 3). It is assumed that μ is a Gaussian random variable with mean equalling (a + b)/2, representing a bias from the nominal μ used to determine the gains, and standard deviation of (b - a)/2. In both cases, the root distributions are skewed and/or multi-modal, and each of the branches has a pronounced peak. Few roots lie near breakaway points, but rather accumulate nearer to the transmission zero or infinity. Figure 3a shows three of the five possible unstable branches, while for the higher gain, only two branches reach instability. Figure 4 indicates that the resulting $Pr(\Sigma)$ are non-Gaussian. The corresponding probabilities of instability $\mathbb{P}$ are 0.48 and 0.33, indicating that the resulting distributions are dissimilar, even though the standard deviations were equally scaled for each case. (When the Case 1 standard deviation is used with Case 2's gains, $\mathbb{P}$ climbs to 0.96.)

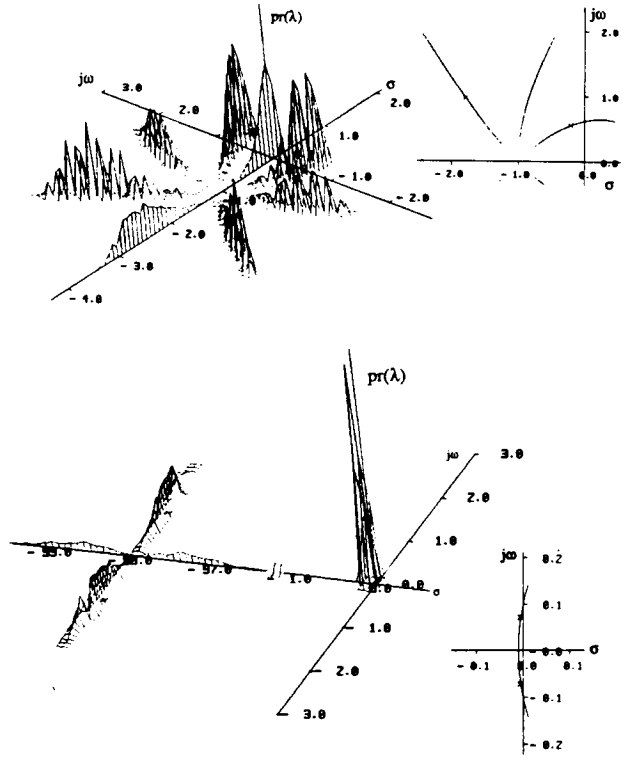


Figure 3. Stochastic root loci for the Doyle LQG counterexample. Gaussian parameter uncertainty with mean = (a + b)/2 and standard deviation = (a - b)/2.
a) c = k = 4, b) c = k = 100

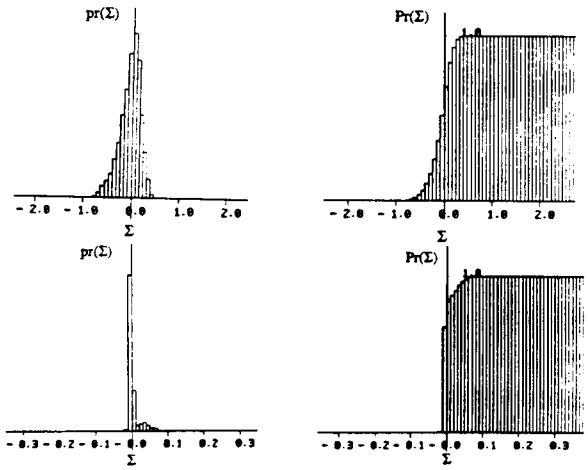


Figure 4. Histograms and cumulative probability distributions for the Doyle LQG counterexample. Gaussian parameter uncertainty with mean = (a + b)/2 and standard deviation = (a - b)/2.
a) c = k = 4, b) c = k = 100

Now consider two similar cases in which μ is a random variable with uniform probability in (a,b). For Case 1, Figures 5 and 6 illustrate how the stochastic root locus and probability distributions are bounded in comparison with Fig. 3a and 4a. In this example, the bounds given by Routh's criterion are not the actual stability bounds, and the probability of instability is non-zero. For Cases 1 and 2, the probabilities of instability $\mathbb{P}$ decrease to 0.27 and 0.01, respectively, as some unstable values associated with the tails of the μ distribution have been eliminated. Naturally, if μ had been uniformly distributed just inside the actual stability boundaries ($0.9243 < \mu < 1.0625$, for c = 4), $\mathbb{P}$ would be zero.

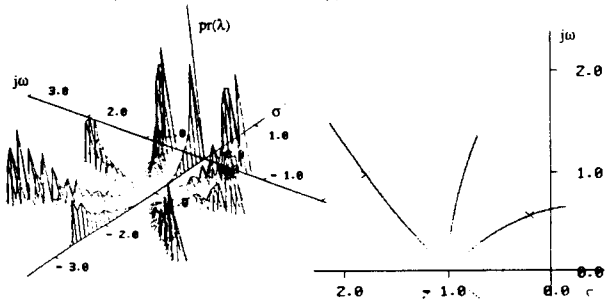


Figure 5. Stochastic root locus for the Doyle LQG counterexample, c = k = 4. Parameter uniformly distributed in (a,b) = (0.875,1.0625).

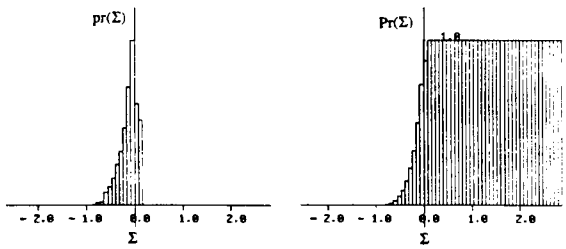


Figure 6. Histogram and cumulative probability distribution for the Doyle LQG counterexample, c = k = 4. Parameter uniformly distributed in (a,b) = (0.875, 1.0625)

Using Loop Transfer Recovery (LQG/LTR) [17], linear-quadratic (LQ) robustness can be fully recovered. Recovery as a function of design parameter ν ($W=\nu W_o$) for Case 1 is illustrated in Fig. 7 through both singular-value analysis and the stochastic root locus. The LQG return difference function in this case is a scalar, and the singular value is identically the return difference function:

$$1+a(s) = \mathbf{I}+\mathbf{C}[s\mathbf{I}-(\mathbf{F}-\mathbf{GC}-\mathbf{KH})]^{-1}\mathbf{KH}[s\mathbf{I}-\mathbf{F}]^{-1}\mathbf{G} \qquad (24$$

The original LQ stability margins are not fully recovered until $\nu > 10{,}000$ (Fig 7a). Figure 7b illustrates the mechanism of recovery: increasing ν pushes two eigenvalues to higher frequencies and decreases the variation in the two roots near the origin. Based upon 3,500 evaluations and a Gaussian μ variation, the present analysis estimates $\mathbb{P}$ to be zero when $\nu \geq 100$.

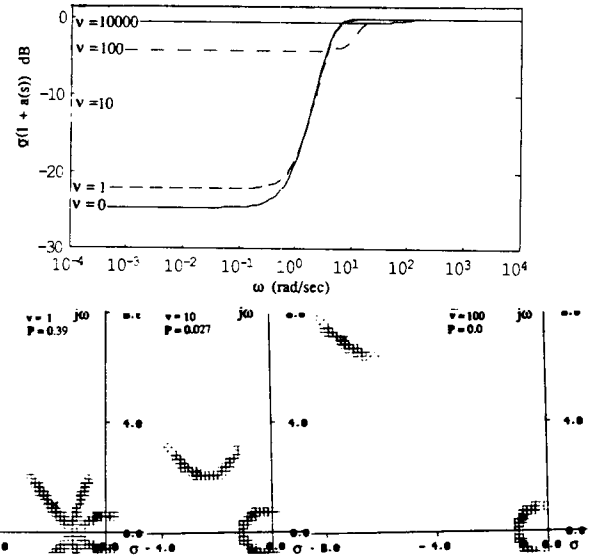


Figure 7. LQG/LTR applied to the Doyle counterexample.
a) Unstructured-singular-value analysis.
b) Stochastic root loci

COMPUTATIONAL ISSUES

The validity of the Monte Carlo analysis is dependent on the number of eigenvalues computed, the number of varying parameters, their probability distributions, and required confidence levels. The number of evaluations required can be related to the number of varying parameters by considering uniform probability distributions. Quantized uniform distributions approximate continuous uniform distributions, approaching them in the limit as the numbers of discrete parameter values go to infinity. Given n Monte Carlo evaluations of a system with r continuous uniform parameters, the result is, at best, equivalent to results computed deterministically for a system with r uniform parameters quantized in w levels, where $w = n^{1/r}$. Conversely, the number of evaluations should be of $O(\alpha w^r)$, where w is an acceptable level of parameter quantization , and $\alpha >> 1$. Note that in a 10-parameter case, direct equivalence to

ten parameter quantization levels requires over 10 billion evaluations, while 10,000 evaluations yield results that are equivalent to a quantization level less than three. Work remains to be done in associating small-sample evaluations with confidence levels of the histograms and the resulting probability of instability.

If σ_{max} is monotonic in the individual elements of $\mathbf{p}$, then evaluation results for the binary probability distribution denoted by $(\mathbf{p}_{min},\mathbf{p}_{max})$ circumscribe results obtained for continuous or quantized distributions with the same limits. In this case, a conservative estimate of $\mathbb{P}$ is provided by the associated 2^r deterministic evaluations based on binary distributions.

Because each Monte Carlo evaluation can be calculated independently, determining stochastic robustness is a task well-suited to parallel computation. Eigenvalue computation speed is linear in the number of processors, and results from separate processors need be consolidated only at the final stage of display.

CONCLUSIONS

Stochastic robustness offers a rigorous yet straightforward alternative to current metrics for control system robustness that is simple to compute and is unfettered by normally difficult problem statements, such as non-Gaussian statistics, products of parameter variations, and structured uncertainty. The approach answers the question, "How likely is the closed-loop system to fail, given limits of parameter uncertainty?" It makes good use of modern computational and graphic tools, and it is easily related to practical design considerations. The principal difficulty in applying this method to controlled systems is that it is computationally intensive; however, requirements are well within the capabilities of existing computers. The principal advantage of the approach is that it is easily implemented, and results have direct bearing on engineering objectives.

ACKNOWLEDGMENT

This research has been sponsored by the FAA and the NASA Langley Research Center under Grant No. NGL 31-001-252.

REFERENCES

1) Dorato, P., ed., *Robust Control*, IEEE Press, New York, 1987.

2) Sandell, N.R., Jr., ed., "Recent Developments in the Robustness Theory of Multivariable Systems," Office of Naval Research Rep. No. ONR-CR215-271-1F, Aug 1979.

3) Lehtomaki, N. A., Sandell, N. R. Jr, and Athans, M., "Robustness Results in Linear-Quadratic-Gaussian Based Multivariable Control Designs," IEEE *Transactions on Automatic Control*, Vol. AC-26, No. 1, pp. 75-93, Feb 1981.

4) Doyle, J. C., "Analysis of Feedback Systems with Structured Uncertainties," IEE *Proceedings*, Vol. 129, Part D, No. 6, pp. 242-250, Nov 1982.

5) Tahk, M. and Speyer, J.L., "Modeling of Parameter Variations and Asymptotic LQG Synthesis," IEEE *Transactions on Automatic Control*, Vol. AC-32, No. 9, pp. 793-801, Sept 1987.

6) Yedavalli, R.K., and Liang, Z., "Reduced Conservatism in Stability Robustness Bounds by State Transformation," IEEE *Transactions on Automatic Control*, Vol. AC-31, No. 9, pp. 863-866, Sept 1986.

7) Horowitz, I., "Quantitative Feedback Theory," IEE *Proceedings*, Vol. 129, Part D, No. 6, pp. 215-226, Nov 1982.

8) Kushner, H.J., *Stochastic Stability and Control*, Academic Press, New York, 1967.

9) Stengel, R.F., "Some Effects of Parameter Variations on the Lateral-Directional Stability of Aircraft," AIAA *Journal of Guidance and Control*, Vol. 3, No. 2, pp. 124-131, Apr 1980.

10) Stengel, R.F. *Stochastic Optimal Control: Theory and Application*, John Wiley and Sons, New York, 1986.

11) Siljak, D.D., *Nonlinear Systems, The Parameter Analysis and Design*, J. Wiley & Sons, New York, 1969.

12) Ackermann, J., "Parameter Space Design of Robust Control Systems," IEEE *Transactions on Automatic Control*, Vol. AC-25, No. 6, pp. 1058-1072, Dec 1980.

13) Putz, P., and Wozny, M.J., "A New Computer Graphics Approach to Parameter Space Design of Control Systems," IEEE *Transactions on Automatic Control*, Vol. AC-32, No. 4, pp. 1058-1072, Apr 1987.

14) Boyd, S., Balakrishnan, V., Barratt, C., Khraishi, N., Li, X., Meyer, D., and Norman, S., "A New CAD Method and Associated Architectures for Linear Controllers," IEEE *Transactions on Automatic Control*, Vol. AC-33, No. 3, pp. 268-283, Mar 1988.

15) Evans, W.R., "Graphical Analysis of Control Systems," *Transactions of the American Institute of Electrical Engineers*, Vol. 67, pp. 547-551, 1948.

16) Doyle, J.C., "Guaranteed Margins for LQG Regulators," IEEE *Transactions on Automatic Control*, Vol. AC-23, No. 4, pp. 756-757, Aug 1978.

17) Doyle, J.C., and Stein, G., "Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis," IEEE *Transactions on Automatic Control*, Vol. AC-26, No. 1, pp. 4-16, Feb 1981.

NASA
National Aeronautics and Space Administration

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CP-3063		

4. Title and Subtitle	5. Report Date
Joint University Program for Air Transportation Research 1988-1989	March 1990
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Frederick R. Morrell, Compiler	L-16740
	10. Work Unit No.
	505-66-01-02
9. Performing Organization Name and Address	11. Contract or Grant No.
NASA Langley Research Center Hampton, VA 23665-5225	
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546 and Federal Aviation Administration Washington, DC 20546	Conference Publication
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This report summarizes the research conducted during 1988-1989 under the NASA/FAA-sponsored Joint University Program for Air Transportation Research. The Joint University Program is a coordinated set of three grants sponsored by NASA Langley Research Center and the Federal Aviation Administration, one each with the Massachusetts Institute of Technology, Ohio University, and Princeton University. Completed works, status reports, and annotated bibliographies are presented for research topics, which include computer science, guidance and control theory and practice, aircraft performance, flight dynamics, and applied experimental psychology. An overview of the year's activities for each university is also presented.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Avionics Low-frequency terrestrial navigation Aircraft guidance, navigation, and control Computer science	Unclassified - Unlimited Subject Category 01

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	213	A10

NASA FORM 1626 OCT 86

For sale by the National Technical Information Service, Springfield, Virginia 22161-2171 NASA-Langley, 1990